IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re application of:	Confirmation No: 2134
FOSTER *et al.*	Art Unit: 1654
Appl. No.: 09/937,484	Examiner: Audet, Maury A.
§ 371 Date: January 23, 2002	Atty. Docket: 1581.0870000/TJS/LDB
For: **Use of a Lectin or Conjugates for Modulation of C-Fibre Activity**	

Declaration Under 37 C.F.R. § 1.132 of Dr. John Chaddock

Assistant Commissioner for Patents
Washington, D.C. 20231

Sir:

I, the undersigned, Dr. John Chaddock, do hereby solemnly and sincerely declare that:

1. I am a named inventor of the subject matter described and claimed in U.S. Patent Application No. 09/937,484.

2. I am currently Head of Molecular Biology at Syntaxin, Ltd. in the United Kingdom. As evidenced by the abridged version of my *curriculum vitae* (attached), I have been actively undertaking research in the field of lectin conjugates for the last 18 years. I am an expert in the field of modulating C-fibre neuron activity for therapeutic purposes.

3. I am familiar with the Office Action dated 29 November 2005 for the above-captioned case.

4. Lectin compounds are a well-recognized group of structurally and functionally related molecules that share common structural features. As a first structural feature, lectins possess a highly conserved binding site triad of amino acids, which is known as the "Asp-Gly-Asn triad". The conserved Asp-Gly-Asn triad is disclosed, for example, in Svensson *et al. J. Mol. Biol. 321*: 69-83 (2002) (enclosed herein as Exhibit A), and in Loris *et al. J. Mol. Biol. 335*: 1227-1240 (2004) (enclosed herein as Exhibit B).

5. As a second structural feature, the lectins possess a "lectin fold", which consists primarily of three β-sheets:

- a "flat" six-membered "back" β-sheet;
- a small "top" β-sheet; and
- a curved, seven-stranded "front" β-sheet.

The "lectin fold" is described in detail in Chandra *et al. Prot. Engin. 14*: 857-866 (2001) (enclosed herein as Exhibit C), and in Turton *et al. Glycobiology 14*: 923-929 (2004) (enclosed herein as Exhibit D). The "lectin fold" facilitates presentation of a substrate (sugar) to the lectin in a concave region of the "front" β-sheet. The highly conserved amino acid triad consisting of Asp, Asn and Gly is found within the "lectin fold."

6. The primary sequence of *Erythrina cristagalli* lectin (ECL) is disclosed in Figure 2 of Exhibit A. Alignment of the primary sequence of ECL with the sequence of the *Erythrina corallodendron* lectin (ECorL) in the same Figure 2 of Exhibit A shows that ECL shares 96% sequence homology with ECorL, and reveals 100% identity in the conserved Asp-Gly-Asn triad: Asp89, Gly107 and Asn 133.

7. The primary sequence of the first 244 amino acids of the *Erythrina corallodendron* lectin (ECorL) was first disclosed in Adar *et al. FEBS Lett. 257*: 81-85 (1989) (enclosed herein as Exhibit E). Comparison of the amino acid sequence of ECorL with that of other legume lectins

reveals a high degree of homology overall and 100% identity in the conserved Asp-Gly-Asn triad. *See* Figure 3 in Exhibit E. This highly conserved lectin binding site "Asp-Gly-Asn triad" is well known to those skilled in the art.

8. It is understood by researchers in the field of lectins that the Asp-Gly-Asn triad and the lectin fold are highly conserved distinguishing features of all lectins.

In this regard, Loris R. *et al., Proteins, 1994 Dec; 20(4):330-346* (enclosed herein as Exhibit F) describes monosaccharide recognition by two forms of lentil lectin, via a highly conserved triad of residues, namely Asp 81, Gly 99 and Asn 125. Thomas, C.J. and Surolia A., *Biochem. Biophys. Res. Commun 2000 Feb 16; 268* (enclosed herein as Exhibit G) confirms that an invariant triad of residues Asp 87, Gly 105 and Asn 137 coordinates recognition of L-fucose by fucose-binding legume lectins. These references confirm that the "Asp-Gly-Asn triad" is essential for ligand recognition in a variety of lectins whose crystal structures are known.

Rao VS. *et al., J. Biomol. Struct. Dyn 1998 Apr 15(5):853-860* (enclosed herein as Exhibit H) confirms, with reference to soybean agglutinin, that the invariant "Asp-Gly-Asn triad" is essential for binding carbohydrate. Together with an aromatic residue (Phe or Tyr), these three invariant residues provide the basic frame for the sugar to bind. This is confirmed by Rao VS. *et al., Int. J. Macromol. 1998 Nov; 23(4):295-307* (enclosed herein as Exhibit I), which reviews the sugar binding sites of Erythrina corallodendron (EcorL), peanut lectin (PNA), Lathyrus ochrus (LOLI) and pea lectin (PSL) and confirms that the invariant residue Asp (from loop A), the invariant residue Asn (from loop C) and the invariant residue Gly (from loop B) are required for a tight interaction with the sugar molecule.

8. I further state that all statements made on my own knowledge are true and that all statements made on information and belief are believed to be true and further that willful false

statements and the like are punishable by fine or imprisonment or both, under Section 1001 of Title 18 of the U.S. Code and that such willful false statements may jeopardize the validity of the application or any patent issuing thereupon.

Date: 21st April 2006

John Chaddock, Ph.D.

515090.1

BIOGRAPHICAL SKETCH

Provide the following information for the key personnel in the order listed for Form Page 2.
Follow this format for each person. **DO NOT EXCEED FOUR PAGES.**

NAME	POSITION TITLE
Chaddock, John Andrew	Head of Molecular Biology

EDUCATION/TRAINING *(Begin with baccalaureate or other initial professional education, such as nursing, and include postdoctoral training.)*

INSTITUTION AND LOCATION	DEGREE *(if applicable)*	YEAR(s)	FIELD OF STUDY
University of Durham, UK	BSc.	1988	Molecular Biology / Biochemistry
University of Warwick, UK	Ph.D.	1992	Protein Biochemistry
Open University, UK	Cert. Mgmt	2003	Management

NOTE: The Biographical Sketch may not exceed four pages. Items A and B (together) may not exceed two of the four-page limit. Follow the formats and instructions on the attached sample.

A. Positions and Honors. List in chronological order previous positions, concluding with your present position. List any honors. Include present membership on any Federal Government public advisory committee.

Positions and Employment

1992-1995 Postdoctoral Research, University of Warwick, UK
1995-1996 Postdoctoral Research, University of Warwick, UK
1996-2001 Scientist, Centre for Applied Microbiology & Research, Salisbury, UK
2002-2005 Senior Scientist, Health Protection Agency, Centre for Applied Microbiology & Research, Salisbury, UK
2005-2006 Head of Molecular Biology, Syntaxin Limited, Salisbury, UK

Other Experience

MSc Examiner in the field of protein toxin biochemistry
Co-supervisor of Ph.D. studentship at the University of Bath
Visiting lecturer at University of Bath

Professional Memberships

1988-present Member of Biochemical Society, UK
2002-2004 Member of International Association for the Study of Pain

B. Selected peer-reviewed publications (in chronological order).

1. Wales, R. Chaddock, J. A., Roberts, L. M. & Lord, J. M. (1992) Addition of an ER retention signal to the ricin A-chain increases the cytotoxicity of the holotoxin. *Exp. Cell Res.*, 203, 1-4.
2. Wales, R., Chaddock, J. A., Corben, E. B., Taylor, S. C., Roberts, L. M., Hartley, M. R. & Lord, J. M. (1993) Mutational analysis and possible applications of ribosome-inactivating proteins. In Beadle, D. J., Bishop, D. H. L., Copping, L. G., Dixon, G. K. & Hollomon, D. W. (eds.) BCPC Monograph No. 55: Opportunities for Molecular Biology in Crop Protection, pp99-111.
3. Chaddock, J. A. & Roberts, L. M. (1993) Mutagenesis and kinetic analysis of the active site Glu177 of ricin A-chain. *Protein Eng.*, 6, 425-431.
4. Chaddock, J. A., Lord, J. M., Hartley, M. R. & Roberts, L. M. (1994) Pokeweed antiviral protein (PAP) mutants which permit *E. coli* growth do not eliminate catalytic activity towards prokaryotic ribosomes. *Nucleic Acids Res.*, 203, 1536-1540.
5. Chaddock, J. A., Roberts, L. M., Jungnickel, B. & Lord, J. M. (1995) A hydrophobic region of ricin A-chain which may have a role in membrane translocation can function as an efficient non cleaved signal peptide. *Biochem. Biophys. Res. Commun.*, 217, 68-73.

6. Walker, D., Chaddock, A. M., Chaddock, J. A., Roberts, L. M., Lord, J. M. & Robinson, C. (1996) Ricin A-chain fused to a chloroplast-targeting signal is unfolded on the chloroplast surface prior to import across the envelope membrane. *J. Biol. Chem.*, 271, 4082-4085.
7. Chaddock, J. A., Monzingo, A. F., Robertus, J. D., Lord, J. M. & Roberts, L. M. (1996) Major structural differences between pokeweed antiviral protein and ricin A-chain do not account for their differing ribosome specificity. *Eur. J. Biochem.*, 235, 159-166.
8. Hartley, M. R., Chaddock, J. A. & Bonness, M. S. (1996) The structure and function of ribosome-inactivating proteins. *Trends Plant Sci.*, 1, 254-260.
9. Zhan, J., de Sousa, M., Chaddock, J. A., Roberts, L. M. & Lord, J. M. (1997) Restoration of lectin activity to a non-glycosylated ricin B chain mutant by the introduction of a novel N-glycosylation site. *FEBS Lett.*, 407, 271-274.
10. Chaddock, J. A., Purkiss, J. R., Friis, L. M., Broadbridge, J. D., Duggan, M. J., Fooks, S. J., Shone, C. C., Quinn, C. P & Foster, K. A. (2000) Inhibition of vesicular secretion in both neuronal and nonneuronal cells by a retargeted endopeptidase derivative of *Clostridium botulinum* neurotoxin type A. *Infect. Immun.*, 68, 2587-2593.
11. Chaddock, J. A., Purkiss, J. R., Duggan, M. J., Quinn, C. P., Shone, C. C. & Foster, K. A. (2000) A conjugate composed of nerve growth factor coupled to a non-toxic derivative of *Clostridium botulinum* neurotoxin type A can inhibit neurotransmitter release *in vitro*. *Growth Factors*, 18(2), 147-155.
12. Chaddock, J. A. and Melling, J. (2002) *Clostridium botulinum* and associated neurotoxins. Chapter 55. *Molecular Medical Microbiology*. Ed Max Sussman, Academic Press, London. 1141-1152.
13. Chaddock, JA., Herbert, MH., Ling, R, Alexander, FCG., Fooks, SJ., Revell, D, Quinn, CP., Shone, CC. & Foster, KA. (2002) Expression and purification of catalytically active, non-toxic derivatives of *Clostridium botulinum* toxin type A. Prot. Express. Purif., 25, 219-228
14. Turton, K., Chaddock, JA & Acharya, KR. Botulinum neurotoxins: structure, function and medicine. TiBS 2002, 27(11), 552-558.
15. Duggan, MJ, Quinn, CP, Chaddock, JA, Purkiss, JR, Alexander, FCG, Doward, S, Fooks, SJ., Friis, L, Hall, Y, Kirby, ER, Leeds, NJ, Moulsdale, HJ, Dickenson, A, Green, GM., Rahman, W., Suzuki, Rie, Shone, CC and Foster, KA. Inhibition of release of neurotransmitters from rat dorsal root ganglia by a novel conjugate of a *Clostridium botulinum* toxin A endopeptidase fragment and *Erythrina cristagalli* lectin. J. Biol. Chem. 2002, 277(38), 34846-34852.
16. Stancombe, PR, Alexander, FCG, Ling, RJ, Matheson, MA, Shone CC & Chaddock, JA. Isolation of the gene and large-scale expression and purification of recombinant *Erythrina cristagalli* lectin. Protein Expression & Purification 2003, 30, 283-292.
17. Chaddock, JA, Duggan, MJ, Quinn, CP, Purkiss, JR, Alexander, FCG, Doward, S, Fooks, SJ., Friis, L, Hall, Y, Kirby, ER, Leeds, NJ, Moulsdale, HJ, Dickenson, A, Green, GM., Rahman, W., Suzuki, Rie, Shone, CC and Foster, KA. Retargeted clostridial endopeptidases: Inhibition of nociceptive neurotransmitter release *in vitro*, and antinociceptive activity in *in vivo* models of pain. Movement Disorders 2004, 19(S8), S42-S47.
18. Hall, YHJ, Chaddock, JA, Moulsdale, HJ, Kirby, ER, Alexander, FCG, Marks, JD and Foster, KA. Use of a new in vitro botulinum neurotoxin antibody detection assay to assess novel vaccine candidates. J. Immunological. Methods 2004, 288 (1-2), 55-60
19. Turton, K., Natesh, R., Thiyagarajan, N., Chaddock, JA & Acharya, KR. 2004, Crystal structure of Erythrina cristagalli lectin with bound N-linked oligosaccharide and lactose. Accepted by Glycobiology
20. Sutton, JM, Wayne, J, Scott-Tucker, A, O'Brien, SM, Marks, PMH, Alexander, FCG, Shone, CC and Chaddock, JA. Preparation of specifically activatable endopeptidase derivatives of Clostridium botulinum toxins type A, B and C and their applications. Accepted by Protein Expression & Purification
21. Chaddock, JA & Marks, PMH. Clostridial neurotoxins: structure-function led design of new therapeutics. Cellular and Molecular Life Sciences, 2006, 63, 540-551.

doi:10.1016/S0022-2836(02)00564-5 available online at http://www.idealibrary.com on IDEAL J. Mol. Biol. (2002) 321, 69–83

JMB

High-resolution Crystal Structures of *Erythrina cristagalli* Lectin in Complex with Lactose and 2′-α-L-Fucosyllactose and Correlation with Thermodynamic Binding Data

Cecilia Svensson[1], Susann Teneberg[2], Carol L. Nilsson[2] Anders Kjellberg[2], Frederick P. Schwarz[3], Nathan Sharon[4] and Ute Krengel[1]*

[1]*Department of Molecular Biotechnology and Center for Structural Biology, Chalmers University of Technology, P. O. Box 462, SE-405-30 Goteborg Sweden*

[2]*Department of Medical Biochemistry and Center for Structural Biology, Göteborg University, P. O. Box 440 SE-405-30 Goteborg, Sweden*

[3]*Center for Advanced Research in Biotechnology, Rockville MD 20850, USA*

[4]*Department of Biological Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel*

**Corresponding author*

The primary sequence of *Erythrina cristagalli* lectin (ECL) was mapped by mass spectrometry, and the crystal structures of the lectin in complex with lactose and 2′-α-L-fucosyllactose were determined at 1.6 Å and 1.7 Å resolution, respectively. The two complexes were compared with the crystal structure of the closely related *Erythrina corallodendron* lectin (ECorL) in complex with lactose, with the crystal structure of the *Ulex europaeus* lectin II in complex with 2′-α-L-fucosyllactose, and with two modeled complexes of ECorL with 2′-α-L-fucosyl-*N*-acetyllactosamine. The molecular models are very similar to the crystal structure of ECL in complex with 2′-α-L-fucosyllactose with respect to the overall mode of binding, with the L-fucose fitting snugly into the cavity surrounded by Tyr106, Tyr108, Trp135 and Pro134 adjoining the primary combining site of the lectin. Marked differences were however noted between the models and the experimental structure in the network of hydrogen bonds and hydrophobic interactions holding the L-fucose in the combining site of the lectin, pointing to limitations of the modeling approach. In addition to the structural characterization of the ECL complexes, an effort was undertaken to correlate the structural data with thermodynamic data obtained from microcalorimetry, revealing the importance of the water network in the lectin combining site for carbohydrate binding.

Keywords: crystal structure; glycobiology; lectin; protein–carbohydrate interactions; structure/function

Introduction

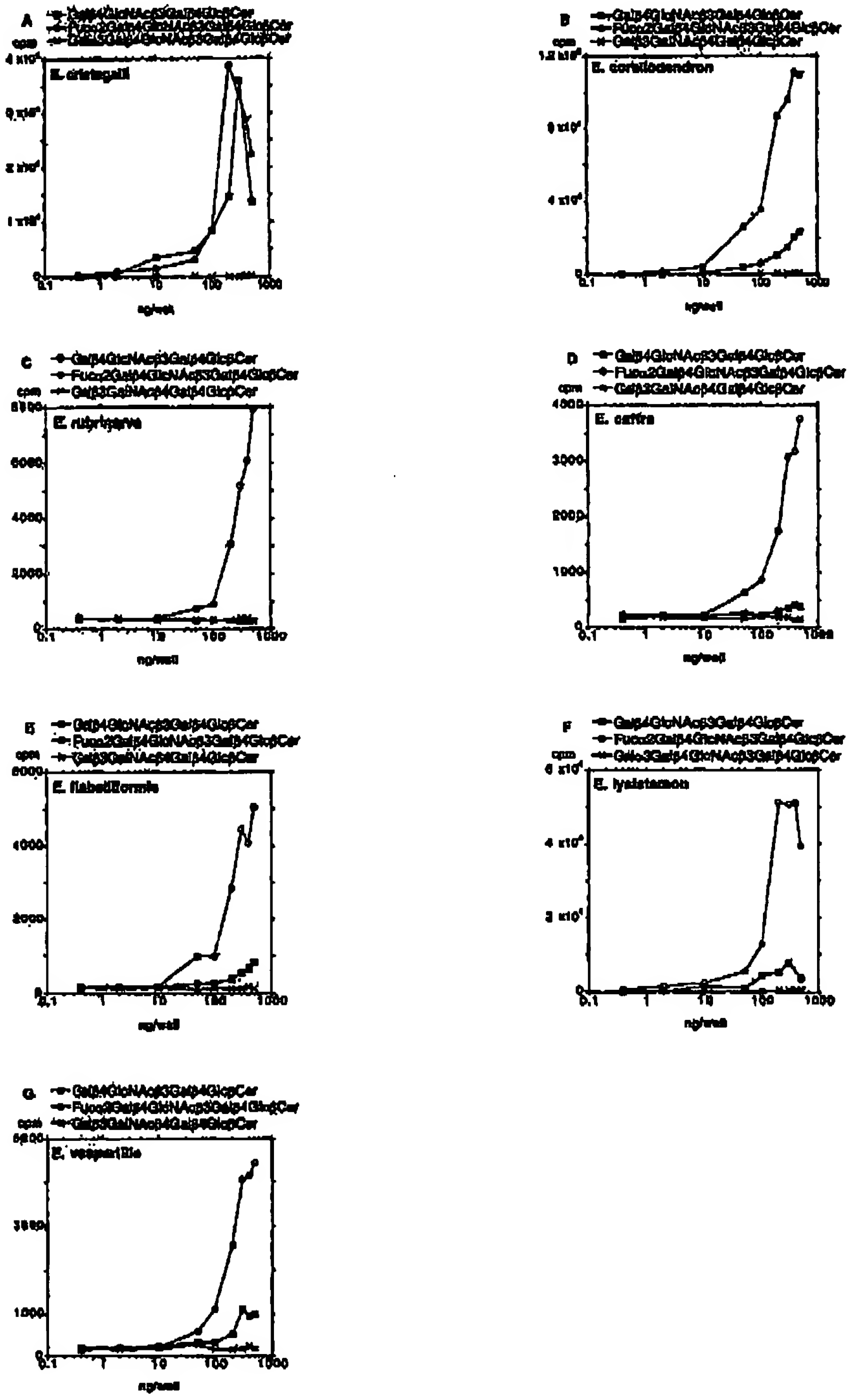
Specific recognition of carbohydrates by proteins lies at the heart of many biological processes, ranging from cell–cell interaction and adhesion of infectious agents to host cells, cellular signaling and differentiation, malignancy and metastasis, to fertilization and immune response.[1,2] Understanding the roles of carbohydrates in these processes and how they interact with proteins is expected to have a large impact on the development of new treatments against many human diseases.

Legume lectins are especially well suited as a model system to study the molecular basis of protein–carbohydrate recognition because they are structurally similar and yet their specificity is diverse.[3–5] By now, the three-dimensional structures of 20 legume lectins have been solved by high-resolution X-ray crystallography, both in free form and in complex with a variety of carbohydrate ligands†.

Abbreviations used: ECL, *Erythrina cristagalli* lectin; ECorL, *Erythrina corallodendron* lectin; ELLA, enzyme-linked lectin assay; Fuc, L-fucose; fucosyllactose, FucLac, 2′-α-L-fucosyllactose; fucosyl-*N*-acetyllactosamine, 2′-α-L-fucosyl-*N*-acetyllactosamine; Gal, galactose; Glc, glucose; ITC, isothermal titration calorimetry; Lac, lactose; MPD, 2-methyl-2,4-pentanediol; MS, mass spectrometry; rms, root mean square; PBS, phosphate-buffered saline; UEA-II, *Ulex europaeus* lectin II, WBA-I and II, winged bean acidic lectin I and II.

E-mail address of the corresponding author: ute.krengel@molbiotech.chalmers.se

† See 3D Lectin Data Bank on World Wide Web URL: http://www.cermav.cnrs.fr/databank/lectine


A
E. cristagalli
B
E. corallodendron
C
E. rubrinerva
D
E. caffra
E
E. flabelliformis
F
E. lysistemon
G
E. vespertilio
cpm
ng/well

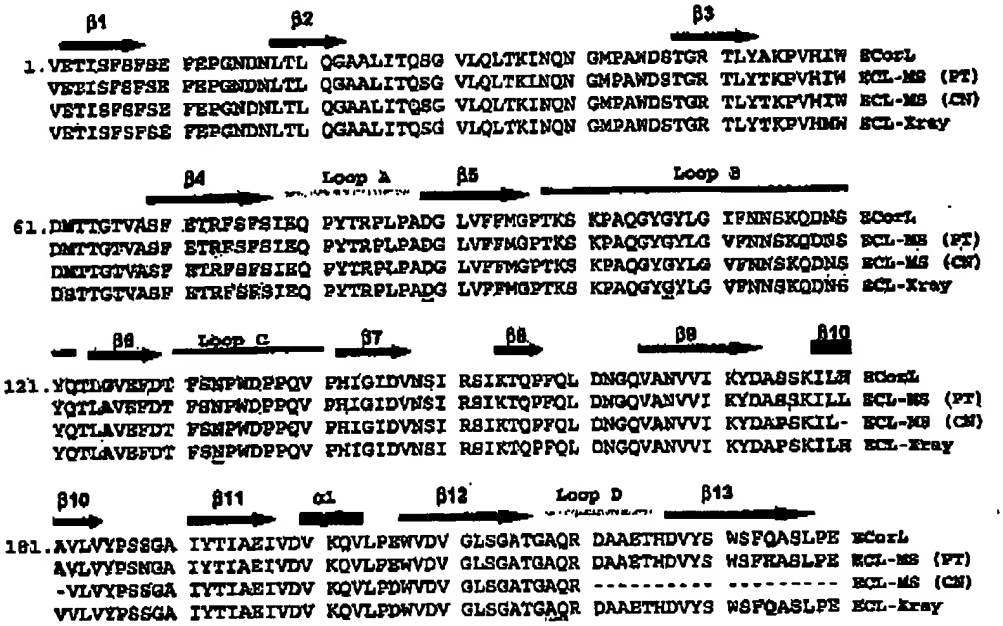


Figure 2. Alignment of the peptide sequences of ECL and ECorL (SWISS-PROT entry P16404). The ECL sequence was derived, both by mass spectrometry (C.N., present study; P. Thibault, personal communication, on ECL from a different source) and from crystallographic electron density maps. Secondary structural elements and important interaction partners of the lactose ligand (the Asp89-Gly107-Asn133 triad and Ala218-Gln219, underlined) are highlighted. Differences between the four sequences are indicated in bold.

One of the most thoroughly studied of these is the Gal/GalNAc specific *Erythrina corallodendron* lectin (ECorL), the best natural ligand of which is the H-2 blood type determinant fucosyl-*N*-acetyllactosamine.[6] The crystal structures of ECorL and of its complexes with several mono- and disaccharides, notably lactose and *N*-acetyllactosamine, have been determined,[7,8] and models have been presented of the complex with fucosyl-*N*-acetyllactosamine.[6,9] In addition, calorimetric data for these as well as other lectin–carbohydrate complexes have become available.[10–14] The relationship between structural and thermodynamic data is currently under discussion.[8,9,14–16]

In the present investigation, we have chosen to study several highly homologous legume lectins from the Erythrina family as a model system for protein–carbohydrate interactions. Particular emphasis is on ECL,[17] which, despite its high homology, exhibits subtle, but distinct differences in carbohydrate binding properties compared to the other members of this family. The aim of the current study was to relate the results from carbohydrate-binding assays and microcalorimetry to structural data from X-ray crystallography and primary sequence analysis in order to characterize the underlying protein–carbohydrate interactions of this system, both in structural and thermodynamic terms.

Results and Discussion

Carbohydrate binding studies

The lectins of the Erythrina family were initially characterized as specific for galactose and *N*-acetylgalactosamine, with a pronounced preference for *N*-acetyllactosamine.[18] However, binding to solid phase-immobilized glycosphingolipids,[8,19] microcalorimetry,[10] and enzyme-linked lectin assay (ELLA)[9] demonstrated that ECL and ECorL interact more strongly with fucosyllactose and fucosyl-*N*-acetyllactosamine. Subsequently, a difference in the carbohydrate binding preferences of the two lectins was demonstrated. While ECL bound to

Figure 1. Binding of ^{125}I-labeled Erythrina lectins to serial dilutions of glycosphingolipids in microtiter wells. The assay was performed as described in Materials and Methods. Data are presented as mean values of triplicate determinations, after subtraction of the (low) background values.

Table 1. Data collection and refinement statistics

Data set	ECL–Lac	ECL–FucLac
PDB code	1GZC	1GZ9
Unit cell		
a, b (Å)	61.8	80.9
c (Å)	126.1	125.9
Space group	$P4_12_12$	$P4_12_12$
Resolution (Å)	1.6 (1.59–1.64)	1.7 (1.70–1.80)
Number of observed reflections	551,905 (38,515)	429,936 (39,733)
Number of unique reflections	58,614 (6065)	46,349 (7167)
R_{merge} (%)	5.3 (60.8)	6.7 (69.3)
Completeness (%)	99.5 (99.4)	99.3 (99.2)
I/σ	20.9 (2.3)	17.5 (2.6)
Redundancy	9.4 (6.4)	9.3 (5.5)
R_{cryst} (%)	20.8 (33.4)	20.8 (33.8)
R_{free} (%)	22.6 (34.3)	23.9 (34.6)
Rmsd bond lengths (Å)	0.018	0.014
Rmsd angles (°)	1.90	1.98
Real space correlation coefficient[a] (%)	92.4	91.6
Ramachandran profile[b]		
Most favorable (%)	88.2	88.2
Additionally allowed	11.3	11.3
Generously allowed	0.5	0.5
Disallowed	0.0	0.0

Values in parentheses correspond to the data in the highest resolution shell, for all data with a signal to noise ratio $\geq -3\sigma$.
[a] Calculated with the program O.[45]
[b] According to the program PROCHECK.[21]

N-acetyllactosamine- and fucosyl-*N*-acetyllactosamine-terminated compounds with similar affinity,[19] the latter compounds were the preferential ligands for EcorL.[6] However, no difference in binding affinities of the two lectins was observed to complex type oligosaccharides and synthetic cluster glycosides, which do not contain a fucose residue.[20] Glycosphingolipid binding experiments with other members of the Erythrina lectin family (*E. rubrinerva, E. vespertilio, E. lysistemon, E. caffra,* and *E. flabelliformis*; Figure 1(c)–(g)) demonstrated that their carbohydrate binding preferences were similar to ECorL (Figure 1(b)), i.e. they preferentially bound to the fucosyl-*N*-acetyllactosamine-terminated compound.

The finding that the carbohydrate binding properties of ECL (Figure 1(a)) differ from those of the other members of the Erythrina lectin family, despite their high homology, prompted us to attempt the crystallographic analysis of ECL, both in complex with lactose and fucosyllactose, in order to structurally characterize the underlying protein–carbohydrate interactions of this system.

Primary sequence

The ECL sequence was obtained by mass spectrometric peptide mapping[21–23] and tandem mass spectrometry[24] of the enzymatically digested protein purchased from Vector Laboratories. The sequence determination was aided by the available sequence of a different ECL isoform (P. Thibault, personal communication). Sequence coverage of 213 of 239 amino acid residues, or 89%, was obtained. Comparison with the electron density obtained from the crystallographic analysis showed excellent agreement (Figure 2), except for residues 59 and 62. Residue 59 is a methionine in the X-ray structure and an isoleucine, or the isobaric leucine according to the mass spectrometry data, while residue 62 was modeled as a serine in the X-ray structure, but is a methionine according to mass spectrometry results. The difference at the former residue most likely reflects the presence or dominance of different isoforms of ECL in the crystal and in solution, respectively, while for the latter difference, it cannot be excluded that it represents an artifact caused by crystal disorder. Overall, the ECL primary sequence is highly similar to that of ECorL, with a sequence identity of approximately 97%. The combining site residues (thus all residues that interact either directly with the carbohydrate ligands or indirectly *via* a water molecule) exhibit 100% sequence identity.

Structural analysis

Accuracy of crystal structures

The crystal structures of the ECL complexes with lactose and fucosyllactose have been determined at 1.6 Å and 1.7 Å, respectively, based on high quality, strong and highly redundant data. At this high resolution it is possible to interpret many details in the electron density maps, including the positions of water molecules and the existence of several alternative conformations for protein side-chains or hydroxyl groups of the carbohydrate ligands.

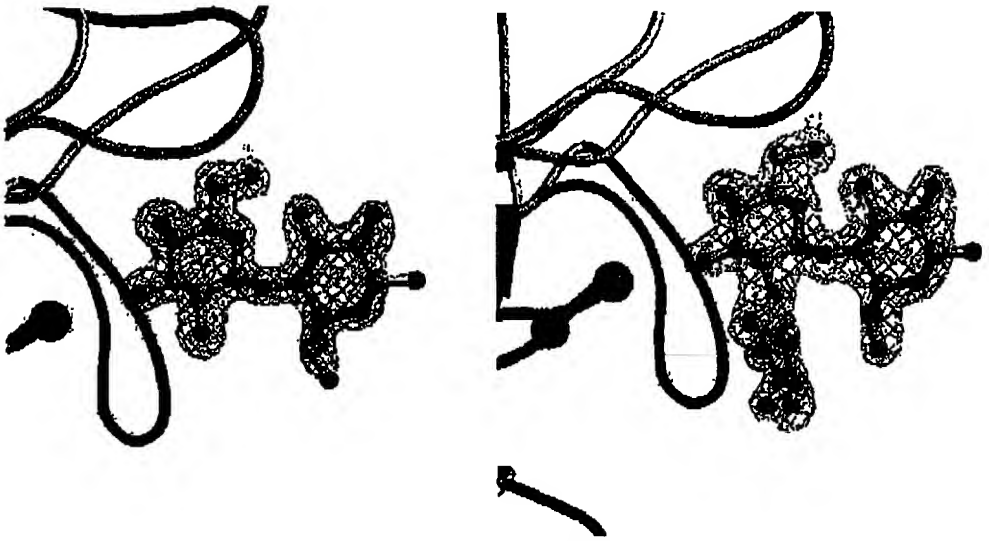


Figure 3. Electron density for lactose (a) and fucosyllactose (b), in the respective ECL complexes. $2F_o - F_c$ simulated-annealing omit maps covering the saccharides, displayed at 1σ.

The final *R*-factors for the ECL structures are 20.8% (R_{free} = 22.6%) and 20.8% (R_{free} = 23.9%) for the lactose and fucosyllactose complexes, respectively. Rms deviations from ideal geometry are 0.018 Å/0.014 Å and 1.90°/1.98° for the respective bond lengths and angles of the two complexes. Finally, the Ramachandran plots[25] are consistent with geometrically well-defined structures (Table 1). All of these values reflect the high quality of the structural models. At the same time, the structures of the ECL complexes are well defined by electron density for essentially all 239 amino acid residues identified by mass spectrometry as well as for the carbohydrate ligand (Figure 3). The only slightly weaker points of the structure are at the 2-fold symmetry axis around His180. The good agreement between the density and the structural model is also reflected in the average real-space correlation coefficients of 92.4% and 91.6%, respectively, for simulated annealing omit maps of the lactose and fucosyllactose complexes.

ECL structure

Tertiary and quaternary structures. As in the case of ECorL and other legume lectins, the ECL fold

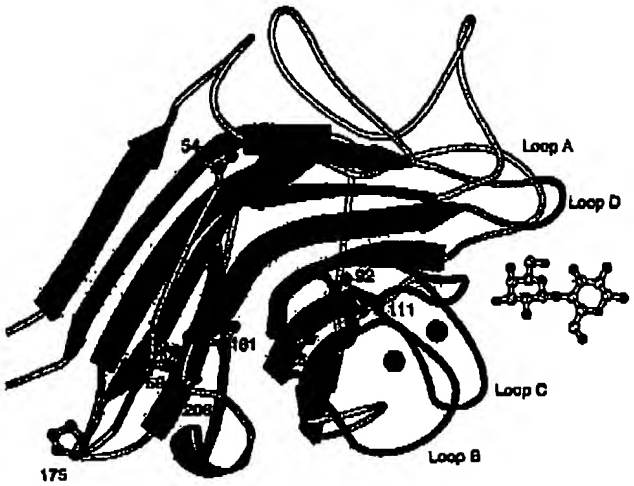


Figure 4. Topology of ECL (complex with lactose). ECL residues that differ from ECorL are shown explicitly. In addition, Val 92, the interaction partner of residues 111 and 125, is highlighted in orange.

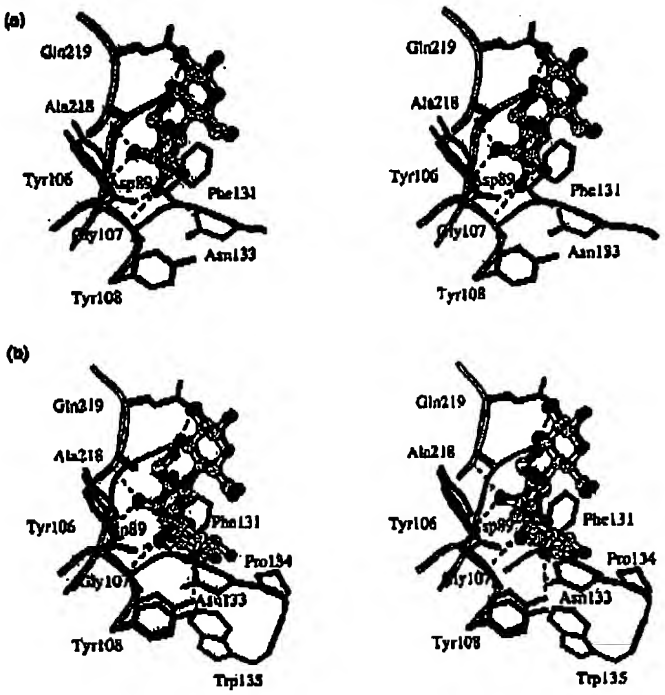


Figure 5 (*legend opposite*)

consists of two stacked antiparallel β-sheets (Figure 4). In addition to the β-sheets, ECL exhibits one α-helical turn between strands β11 and β12, according to analysis using the program DSSP.[26] Similar to ECorL, WBA-I and WBA-II,[27] ECL forms a non-canonical dimer, in which two monomers are arranged back-to-back, as if engaged in a "handshake". The lectin dimer interface is characterized by extensive, mostly hydrophobic interactions between the two monomers. Since ECL crystallized with only one molecule per asymmetric unit, the two protomers are crystallographically equivalent.

Both ECL and ECorL have two *N*-glycosylation sites, at Asn17 and Asn113†. Mass spectrometry analysis of ECL revealed that both residues are linked to a heptasaccharide of the structure (deoxyhexose)$_1$(pentose)$_1$(hexose)$_3$(*N*-acetylhexosamine)$_2$, with partial occupancy at Asn113. At both sites, the electron density for the attached saccharide is rather weak, revealing only the rough positions of one to three sugar residues. For this reason, none of the sugar rings has been modeled. This is in contrast to the situation in ECorL, where the heptasaccharide linked to Asn17 is exceptionally well defined by electron density.[7] A comparison of the crystal environments of ECL and ECorL revealed packing interactions as the origin of these differences. While the *N*-linked saccharide in ECorL is involved in extensive interactions with the Asn17-linked heptasaccharide of a symmetry-related molecule, the same residue in the ECL crystal structure is highly solvent-exposed.

It has been the focus of some debate, whether glycosylation of Asn17 prevents the formation of the canonical dimer.[8,7,27–29] On the basis of the ECL structure, this possibility cannot be ruled out, although the nature of the protein surface involved in dimerisation suggests that the quaternary structure is determined to a significant extent by factors intrinsic to the protein itself, independent of glycosylation.

The carbohydrate-binding site. The ECL combining site is located in a shallow, highly solvent-exposed

† It was noted that glycosylation at Asn113 could affect the carbohydrate binding properties of the lectins under investigation, as the bound heptasaccharide is located on the same side of the lectin as the combining site. Glycosylation of Asn113 could therefore critically influence lectin binding to glycoconjugates and would then be insufficiently modeled by experiments involving only small saccharide analogs like lactose and fucosyllactose. However, previous experiments with recombinant ECorL, which is not glycosylated, show that no differences in binding properties are observed compared to the native lectin,[6] thus confirming the validity of our system.

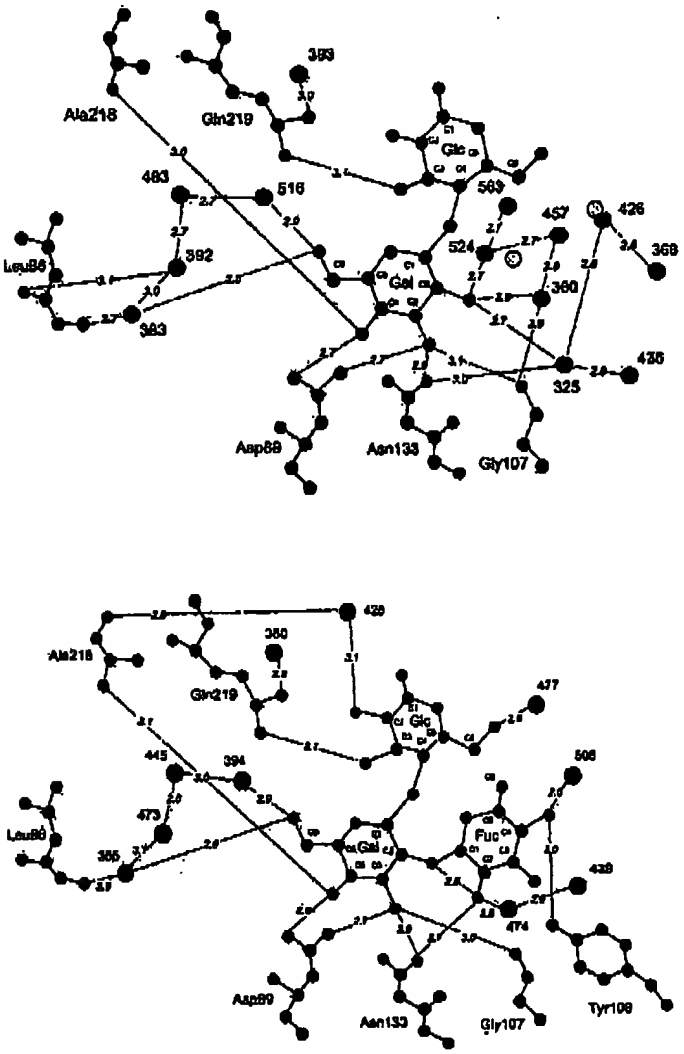


Figure 5. ECL combining site. (a) and (b) Stereo-pictures showing a superposition of the combining sites of (a) ECL (blue) and ECorL (white/red) in complex with lactose, (b) ECL in complex with lactose (blue) and fucosyllactose (white/red), respectively. Colors of the loops correspond to those in Figures 3 and 4. Hydrogen bonds between the saccharide and ECL are indicated. (c) and (d) 2D-plots of the combining sites of ECL in complex with lactose (c) and fucosyllactose (d), respectively. Water molecules of ECorL, which are not present in the ECL combining site, are indicated as yellow circles.

depression on the surface of each protomer. Several highly conserved amino acid residues form the site and provide the contacts necessary for a number of strong interactions with the carbohydrate ligands (see Figure 5). In particular, Asp89, Gly107 and Asn133 of the conserved Asp-Gly-Asn triad create the basis for a strong hydrogen-bonding network. Additional important interaction partners of the carbohydrate ligands are the aromatic ring of Phe131, which provides hydrophobic stacking interactions with galactose C3/4/5, and the main chain carbonyl oxygen of Leu86, which interacts with galactose O6 *via* a structural water molecule. At the periphery of the combining site, Ala218 and Gln219 from the specificity loop D^{20} engage in somewhat weaker hydrogen-bonding interactions with the disaccharide.

The carbohydrate binding activity of legume lectins depends on the presence of two metal ions, a calcium ion and a transition metal ion (usually manganese), which although not involved in directly binding the carbohydrate ligands, help to position the amino acid residues in contact with them. The metal ion coordination in ECL is identical with ECorL, involving four protein ligands and four water molecules. The four metal-coordinated water molecules are conserved in all legume lectins inspected,[31] one of them being essential for

Table 2. *B*-factors for water molecules and saccharide units in the combining sites of the ECorL–Lac, ECL–Lac and ECL–FucLac complexes (structurally equivalent water molecules are placed on the same line)

ECorL–Lac	*B*-factor (Å²)	ECL–Lac	*B*-factor (Å²)	ECL–FucLac	*B*-factor (Å²)
Wat 522	31.4	Wat 383	31.9	Wat 385	27.6
Wat 523	39.6	Wat 392	33.4	Wat 479	29.0
Wat 524	51.9	Wat 516	39.9	Wat 394	34.1
		Wat 483	44.8	Wat 445	47.3
Wat 525	59.2	Wat 325	24.7		
Wat 526	55.2	Wat 436	42.0		
Wat 527	71.3				
Wat 528	53.4	Wat 360	24.7		
Wat 586	44.6	Wat 393	39.2	Wat 380	34.0
Wat 588	60.0			Wat 428	50.8
Wat 626	54.9	Wat 368	27.8	Wat 361	30.9
Wat 627	62.1	Wat 426	38.6		
		Wat 524	40.3		
		Wat 457	38.9		
		Wat 563	45.2		
		Wat 549	41.6	Wat 439	37.6
				Wat 474	37.5
				Wat 477	48.7
				Wat 508	53.1

B-factors (Å²)	ECorL–Lac	ECL–Lac	ECL–FucLac
Overall	32.4	27.1	29.5
Gal	29.8	28.1	27.8
Glc	52.2	40.5	44.6
Lac	41.5	34.3	35.8
Fuc			37.5
FucLac			36.3

stabilizing the unusual Ala88-Asp89 *cis*-peptide bond typical for legume lectins.

ECL and ECorL

Overall comparison. Despite the fact that ECL and ECorL crystallized in different space groups, their structures are almost superimposable, with rms differences between C^{α} atoms of the two lactose complexes being 0.37 Å (0.28 Å if only residues 85–220 are considered). In the combining sites of ECL and ECorL, all direct as well as indirect ligands of lactose *via* water molecules are conserved at identical positions, within the error limits of the coordinates (Figure 5(a) and (c)). This is also true for the metal binding sites. The only noticeable differences involve the orientation of Tyr106 in the second shell around the monosaccharide-binding site, the orientation of the glucose ring of the lactose molecule as well as the positions of a few water molecules (ECL 426, 457, 524, 563, Figure 5(c)). Overall, the water network is somewhat strengthened in ECL, as judged from the larger number of interaction partners, shorter H-bonding distances and lower relative *B*-factors (Table 2). The same observation is valid for the glucose unit of lactose, which is less disordered in the ECL complex, whereas the galactose unit is somewhat more ordered in ECorL (Table 2).

Analysis of amino acid differences. ECL and ECorL exhibit only few differences in their primary sequences. These are A54T, I111V, G125A, S175P, A181V and E206D (written as ECorL#ECL) (Figures 2 and 4). All of these substitutions have a very limited effect on the three-dimensional structure. The only substitutions that are within reach of the carbohydrate-binding site and therefore have the potential to cause the altered binding affinities compared to ECorL, are I111V and G125A. These two residues, although distant in sequence, are very close to each other in space, with their side-chains facing each other (Figure 4). They are part of a van der Waals interaction network involving also Val92, Phe112, Phe94, Val123, and, more peripherally, Leu109 and Ile150, in an internal hydrophobic pocket of the protein. Val92 forms a bridge between the two residues 111 and 125, by engaging in van der Waals interactions with the side-chains of both residues. The two "volume-conserving"[33] substitutions at positions 111 and 125 compensate for each other, with Val92 adapting to the change by a slight rotation of its side-chain. Even though the overall changes induced by these substitutions are small as well, they are nevertheless the strongest candidates for causing long-range effects on carbohydrate binding, due to their location and their unique sequence compared to other Erythrina lectins (N. Sharon *et al.*, unpublished results). In such a scenario, Val92 might mediate these effects, by acting as a handle that is pulled in order to transmit the changes to the combining site. Alternatively, a substitution of residue 111 might affect the combining site more directly, through residues 106–108. Single or double mutants of residues 111 and 125 of ECL or ECorL may shed light on this possibility.

Binding of lactose versus *fucosyllactose*

A superposition of the lactose and fucosyllactose complexes of ECL reveals no major rearrangements in the combining site (Figure 5(b)). In fact, both complexes superimpose extremely well, with rms differences of 0.12 Å for the C^{α} atoms of residues 85–220, respectively (rmsd = 0.15 Å for all atoms). Also the two lactose units superimpose quite well, except for a small difference in the relative orientations of the glucose *versus* the galactose rings and a concomitant difference in relative *B*-factors for the galactose unit (Table 2).

The fucose unit, which is linked to lactose *via* the C2-position of galactose, is positioned in a shallow hydrophobic cavity adjacent to the monosaccharide-binding site, shaped by Pro134, Trp135, Tyr108 and Tyr106. This cavity is connected to the primary site by the side-chains of Asn133 on the one side and Ala218 on the other side. These two residues form a gate of 8.2 Å in width, through which the fucose unit stretches into its binding site. Similar, though not identical positioning of fucose was described in modeling studies of ECorL with fucosyl-*N*-acetyllactosamine.[6,9]

Upon binding of the fucose moiety, seven water molecules are displaced from the ECL combining site, four of which are conserved between ECL and ECorL (compare Table 2). This loss is compensated for to a limited extent, by the addition of three new water molecules, two of which engage in strong hydrogen bonds with the 2-OH and 4-OH groups of the fucose unit (Figure 5(c) and (d)). Two additional hydrogen bonds of the fucose unit are to the protein (Figure 5(d)). The fucose 2-OH is hydrogen bonded to Asn133 $N^{\delta2}$, which also interacts with the 3-OH group of galactose and with the glycosidic oxygen of the Fuca2Gal linkage, while the fucose 4-OH is H-bonded to the hydroxyl group of Tyr108. In order to engage in the latter interaction, the side-chain of Tyr108 is slightly pulled into the combining site, compared to its position in the ECL lactose complex. Interestingly, ELLA studies of a 4-methoxy derivative of fucosyllactose caused a small increase in affinity to ECorL.[9] Our results on the ECL structure in complex with fucosyllactose suggest that this increase in affinity probably results from van der Waals interactions of the 4-O methyl group with the fucose atoms C2/C3/C5 and possibly also with the Tyr108 side-chain. No hydrogen bonding interactions are observed for the fucose 3-OH group in the ECL complex, in agreement with results from ELLA studies for ECorL on a 3-deoxygenated compound.[9]

In addition to the hydrogen bonds described above, the fucose moiety engages in several strong hydrophobic interactions with the lectin, notably its 2-OH with Pro134 C^{β} (3.3 Å). A strong hydrophobic interaction of the fucose 2-hydroxy group has been postulated by Lemieux *et al.* (however, involving Trp135) for ECorL and fucosyl-*N*-acetyllactosamine, on the basis of comparative ELLA studies.[9] Interestingly, a strong decrease rather than increase in binding for the fucose 2-deoxy analog was found in a study on the related lectin WBA-II,[12] which points to differences in the binding interactions of WBA-II compared to ECL and ECorL. Two further strong hydrophobic interactions of fucosyllactose involve the fucose C3/C4/C5, which abut to 3.3–3.8 Å of the glucose 6-OH, and the fucose methyl group at the C6 position. The latter group creates a distinctly nonpolar surface at C4/C5/C6, which has been speculated to be involved in the interaction with one of the hydrophobic residues close to the monosaccharide-binding site.[10] Indeed, the fucose methyl group is positioned close to Tyr106 $C^{\varepsilon2}/C^{\varepsilon2}$.

Additional interactions of the fucose are of a weaker nature. They involve intra-molecular interactions with the glucose C6/O6 as well as intermolecular ones with the hydrophobic residues lining the fucose-binding cavity. Curiously, although three of the residues in this cavity are aromatic (Tyr106, Tyr108 and Trp135), none of them are involved in hydrophobic stacking interactions with the saccharide, which are so typical for protein–carbohydrate complexes.[33,34] Instead, contacts are mainly established through atoms on the rim of the aromatic residues or of Pro134. Compared to the lactose complex, the side-chain of Tyr106 is slightly reoriented, such that it actually moves slightly away from the fucose residue. The side-chain of Trp135 is held in position through a stacking interaction of the 6-ring portion of its indole with Asn133 C^{β}, an interaction also present in the ECL lactose complex.

Comparison of the fucosyllactose complexes of ECL and UEA-II

To date, only one other crystal structure of a legume lectin has been solved in complex with fucosyllactose, namely that of lectin II from *U. europaeus* (UEA-II).[35] UEA-II belongs to the chitobiose specificity group, but exhibits a promiscuous carbohydrate-binding site[35] with highest affinity for fucosyllactose.[36] A comparison of the fucosyllactose complexes of ECL and UEA-II reveals qualitatively similar interactions of the fucose residue, even though the saccharide units have completely different relative orientations. The fucose 2-OH and 4-OH groups are involved in hydrogen bonds with the protein (UEA-II residues Gly106 and Ser104), and van der Waals interactions occur to several aromatic residues as well as to glucose C6/O6. Even the strong hydrophobic interaction of one of the fucose hydroxyls is conserved, however, engaging the fucose 3-OH and Trp138 (corresponding to ECL residue Trp135) instead of fucose 2-OH and Pro134, as in ECL. These interactions, although qualitatively similar, however have a completely different structural basis, both with respect to the conformation of the trisaccharide and regarding the protein architecture of the combining site. The largest difference

Table 3. Thermodynamics–ITC (ECorL/ECL)

Carbohydrate	K_a (×1000 M^{-1})	$-\Delta G$ (kJ/mol)	$-\Delta H$ (kJ/mol)	$-\Delta S$ (J/mol K)
Gal	1.6/0.8 ± 0.1	18.2/16.7 ± 0.2	13.7/22.3 ± 0.1	−15.3/18.8 ± 0.2
Lac	1.9/3.1 ± 0.1[a]	18.8/19.9 ± 0.1	41.2/29.9 ± 1.4	75.4/33.6 ± 1.4
FucLac	3.7/3.2 ± 0.2	20.3/20.0 ± 0.1	18.0/19.7 ± 1.1	−7.7/0.0 ± 0.9

ECorL: data from previous experiments.[10]
[a] Average of the two closest values only.

concerns ECL residue Tyr106 and UEA-II residue Tyr135, which are located on opposite sides of the fucose rings. Neither of these residues have a structural counterpart in the other lectin, with Tyr135 of UEA-II even lying in the solvent region of the ECL carbohydrate complex.

The structure of the UEA-II combining site may in fact resemble more closely the combining site of WBA-II, as judged from the results from a thermodynamic study of the latter lectin.[12] These studies revealed a slight increase in the binding constant of 3-deoxy fucosyl-*N*-acetyllactosamine, while the 3-methoxy analog bound significantly worse, which suggests that the 3-OH group is involved in a strong hydrophobic interaction with the lectin.

Correlation of structural and thermodynamic data

The results of the microcalorimetry experiments are summarized in Table 3, a typical experiment is shown in Figure 6. A comparison of the thermodynamic parameters underlying the binding of galactose, lactose and fucosyllactose, respectively, revealed similar tendencies for ECL and ECorL: While ΔH is more negative for the binding of lactose compared to galactose and fucosyllactose, ΔS is significantly more favorable for complexation with fucosyllactose or galactose than for lactose binding. Whereas the thermodynamic data for the binding of fucosyllactose are very similar, for ECL and ECorL, respectively, titrations with galactose and lactose revealed significant differences, especially in the entropic terms. The difference in binding constants of ECL and ECorL for lactose and fucosyllactose is much smaller than expected from the experiments on solid-phase immobilized glycolipids†. While binding to fucosyllactose is very similar for ECL and ECorL, binding to lactose differs only by at most a factor of two (the factor of two difference in lactose binding affinity is in agreement with previous results from microcalorimetry experiments and hemagglutination inhibition assays[11,20]).

Knowledge of both the structures of several closely related lectin–carbohydrate complexes and the corresponding thermodynamic data should make it possible to correlate some of the differences in thermodynamic parameters with structural differences. While changes in enthalpy (ΔH) are generally interpreted in terms of differences in hydrogen bonding, electrostatics and van der Waals interactions, entropy changes (ΔS) are essentially attributed to differences in hydrophobic interactions and mobility.

The crystal structures of ECL and ECorL suggest that the general tendencies for the thermodynamic parameters for galactose, lactose and fucosyllactose binding correlate well with the differences in the local water network in the lectin combining sites. Upon lactose as compared to galactose binding, the local water network is strengthened, leading to a more favorable enthalpy due to an increased number of hydrogen bonds to the carbohydrate ligand as well as to the protein (compare[8,10]). This gain in enthalpy is coupled with a less favorable entropy term. Binding of the fucose unit goes hand in hand with the release of some of these strongly bound water molecules, which roughly returns the situation to the galactose case. Thus, upon binding of fucosyllactose compared to lactose, the enthalpy does not increase with the area buried, as was suggested by Elgavish & Shaanan,[6] but in fact decreases. While for ECorL, at least the free energy ($-\Delta G$) increases slightly upon binding of fucosyllactose compared to lactose, neither of these values increases (to a significant extent) for ECL.

With respect to the entropic term, the binding of lactose is highly unfavorable, whereas the binding of galactose or fucosyllactose is more favorable. From a structural perspective, the entropic costs of lactose *versus* galactose binding can probably be related to the ordering of the combining site, not the least due to the mentioned recruitment of structural water molecules to this site. Thus, more water molecules enhance the negative binding enthalpy, but also reduce the disorder in thelectin–carbohydrate–solvent system and hence the entropy change is negative. Upon binding of fucosyllactose, this order is not significantly reduced, but the entropic term is favorably influenced by the hydrophobic interactions of the fucose moiety. Most

† The differences in the results from the microtiter well assays and microcalorimetry may be due to the difference in sensitivity of the two methods. As microtiter well assays are rather insensitive and might only show binding above a certain threshold, the apparent differences between the two methods can easily be explained with a threshold corresponding to a binding constant K_a between 2000 M^{-1} and 3000 M^{-1} and therefore do not influence our interpretations of the microcalorimetric data.

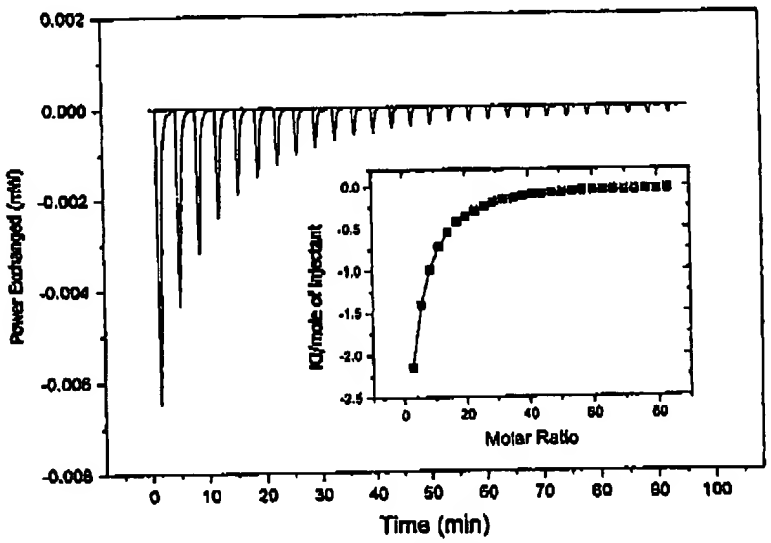


Figure 6. The results of a typical ITC experiment, which consisted of titrating 10 µl aliquots of 12.7 mM lactose into a solution containing 0.0325 mM ECL in 20 mM phosphate and 150 mM sodium chloride buffer at 298 K. The inset shows the binding isotherm fit of the data to a 1:1 binding model.

importantly, the fucose 2-OH engages in a strong hydrophobic interaction with Pro134 C^{β} and the fucose methyl group interacts with Tyr106 $C^{\alpha 2}$.

While the thermodynamic parameters for ECL and ECorL follow similar general trends, a more thorough comparison reveals distinct differences with respect to the binding of galactose and lactose, as mentioned above. In both cases, the differences in binding constants are to a large extent due to different entropic contributions for ECL and ECorL, respectively. It is, however, not straightforward to explain these differences in structural terms, as the differences between the combining sites of ECL and ECorL are minimal. Possibly, the slightly higher relative *B*-factors for the galactose residue of lactose in ECL indicate a somewhat enhanced mobility of the monosaccharide-binding site for this lectin (leading to a more favorable entropic term), which might be caused by the substitutions at residues 111 and 125 and transmitted to the combining site *via* Val92. It would be interesting to probe this system with molecular dynamics simulations similar to the ones performed by Bradbrook *et al.*[16]

Conclusions and Outlook

In the present investigation, the crystal structures of two carbohydrate complexes of ECL have been solved at high resolution, revealing detailed structural information of the interactions of this lectin with lactose and fucosyllactose, respectively. The structures were compared with the crystal structure and with two molecular models of the highly homologous lectin from ECorL with fucosyl-*N*-acetyllactosamine, which was found to exhibit subtle, but distinct differences in binding properties compared to ECL. An effort was undertaken to characterize the protein–carbohydrate interactions of this system in light of the thermodynamic data obtained by microcalorimetry. The results of this analysis point to an important role of the water network in the lectin combining site. We propose that two amino acid substitutions at position 111 and 125 are the cause of the differences in binding affinities between ECL and ECorL. Even though the two substitutions compensate for each other, they cause a slight rotation of the side-chain of the interacting residue Val92, which might transmit the changes to the combining site. The hypothesis proposed can now be tested by creating hybrids of the two lectins that differ in positions 111 and 125.

Materials and Methods

Isolation of Erythrina lectins

The lectins from *E. rubrinerva*, *E. vespertilio*, *E. lysistemon*, *E. caffra*, and *E. flabelliformis* were purified by affinity chromatography on lactose-Sepharose, as described.[18] ECL was purchased from Vector Laboratories Inc. (Burlingame, CA) and Sigma (St. Louis, MO), while ECorL was obtained from Sigma. The lectins were diluted to 1 mg/ml in phosphate-buffered saline (PBS) (pH 7.3), containing 8 mM phosphate buffer (pH 7.3), 0.14 M NaCl and 4 mM KCL. Aliquots of 100 µg were

labeled with ^{125}I, using $Na^{125}I$ (100 μCi/ml; Amersham Pharmacia Biotech, Little Chalfont, UK), according to the IODO-GEN protocol of the manufacturer (Pierce, Rockford, IL). Approximately 5×10^3 cpm/μg protein was obtained.

Reference glycosphingolipids

The glycosphingolipids utilized in the carbohydrate binding studies were isolated by standard methods[17] and characterized by mass spectrometry and proton NMR, as described.[19]

Microtiter well assay

The microtiter well binding assay was performed as described.[19] In brief, serial dilutions (each dilution in triplicate) of pure glycosphingolipids in methanol were applied in microtiter wells (Cooks M24, Nutacon, Holland). When the solvent had evaporated, the wells were blocked for two hours with 200 μl of PBS containing 2% (w/v) bovine serum albumin and 0.1% (w/v) NaN_3 (Sol. 1). Thereafter, 50 μl of radiolabeled Erythrina lectins diluted in Sol. 1 (approximately 2×10^5 cpm/μl), were added per well and incubated for four hours at room temperature. After washing six times with PBS, the wells were cut out and the radioactivity counted in a gamma counter.

Isothermal titration calorimetry (ITC) measurements

The thermodynamic parameters for the binding of galactose, lactose, and fucosyllactose (Dextra Laboratories LTD) to ECL (Sigma†) at 298.15 K were determined by ITC using a Microcal, Inc. VP ITC. The VP ITC consists of a matched pair of sample and reference vessels (1.409 ml) enclosed in an adiabatic enclosure and a rotating stirrer-syringe for titrating aliquots of the ligand solution into the sample vessel. The sample vessel contained 0.02–0.052 mM of the lectin solution, while the reference vessel contained just the 20 mM Na_3PO_4 and 0.15 mM NaCl (pH 7.4) buffer solution. Aliquots (10 μl) of the carbohydrate solution (at concentrations of 1.0–12.7 mM) were titrated four minutes apart into the lectin sample solution until the heat exchanged under the titration peak was reduced by a factor of 5–10. Any contribution of the carbohydrate heat of dilution to the binding enthalpy was determined by titrating the carbohydrate ligand solution directly into the buffer solution in the sample vessel. The heats of dilution were then subtracted from the heats obtained during the titration prior to analysis of the data.

A non-linear, least square minimization software program from Microcal, Inc., Origin 5.0, was used to fit the incremental heat of the ith titration ($\Delta Q(i)$) of the total heat, Q_t, to the total titrant concentration, X_t, with a stoichiometry fixed at $n = 1.0$, according to the following equations

$$Q_t = C_t \Delta H^0_b V\{1 + X_t/C_t + 1/K_b C_t - [(1 + X_t/C_t + 1/K_b C_t)^2 - 4X_t/C_t]^{1/2}\}/2 \quad (1a)$$

$$\Delta Q(i) = Q(i) + dVi/2V[Q(i) + Q(i-1)] - Q(i-1) \quad (1b)$$

where C_t is the total lectin concentration in the sample vessel and V is the volume of the sample vessel, to yield values of K_b and ΔH^0_b. The uncertainties reported for K_b and ΔH^0_b are the standard deviations from the average value, determined from at least two different titration.

Peptide mapping by mass spectrometry

For enzymatic digestion, approximately 10 nmol of ECL (Sigma, St. Louis, USA) was dissolved in 500 μl buffer containing 10 mM *n*-octylglucoside, 0.1 mM $CaCl_2$ and 0.1 M NH_4HCO_3, to which 100 pmol modified trypsin (Promega, Wisconsin, USA) had been added. Digestion of the protein proceeded at 38 °C for ten hours. The mixture was dried in a Speedvac (Savant, Inc.).

Samples were analyzed using a Micromass TofSpecE MALDI-TOF mass spectrometer (Micromass, Manchester, UK) equipped with a pulsed 337 nm nitrogen laser, a delayed extraction ion source and a reflectron. All spectra were acquired in reflectron mode at an accelerating voltage of 20 kV and were the sum of 100 laser shots. External calibration was performed with the monoisotopic masses of angiotensin II and ACTH (18–39). Spectra were analyzed using the MassLynx (Micromass) software in a WindowsNT environment.

Fragment ion data were acquired in an electrospray-quadrupole time-of-flight instrument (QTof, Micromass, Manchester, UK). The protein digest was dissolved in 3 μl of acetonitrile/water (1:1, v/v) containing 0.1% formic acid and sprayed from gold-coated glass capillaries in a nanoflow source. Argon was used as the collision gas. Instrument calibration was performed using fragment ions from Glu-fibrinopeptide B and a fourth-order polynomial fit. MS/MS spectra were post-processed with the MaxEnt3 software (Micromass). The sequencing results are given in Figure 2.

Crystallization

ECL containing 13% lactose by weight (Vector laboratories, Inc. Burlingame, CA) was dissolved in 20 mM Hepes, 100 mM NaCl, to a concentration of 10 mg/ml. Crystallization experiments were performed using the hanging drop vapor diffusion method at room temperature. The first screening (Crystal Screen™ Hampton Research) resulted in small tetragonal pyramidal or bipyramidal crystals from 2 M ammonium sulfate, 0.1 M Tris–HCl (pH 8.0). The crystal size was improved by macroseeding into a pre-equilibrated protein solution containing 1.8 M ammonium sulfate, 0.1 M Tris–HCl (pH 8.0) and 15% glycerol. Crystals grew to a maximum size of 0.2 mm × 0.4 mm × 0.4 mm.

† While the carbohydrate binding assays, mass spectrometry and crystallization experiments were all performed using ECL purchased from Vector Laboratories Inc., the microcalorimetry experiments were done with ECL from Sigma. The reasons for this difference were purely practical, as the lectin preparation from Vector contained large amounts of lactose and was unstable in the absence of carbohydrate. In order to ensure that all experiments in the present study were compatible, we therefore compared the binding pattern of dissolved ECL crystals to the commercially obtained samples of both sources. The obtained curves were very similar to Figure 1(a) (data not shown), thus reassuring us that functional conclusions drawn from the crystal structures of the ECL complexes have a valid basis.

For co-crystallization with fucosyllactose (Dextra Laboratories LTD), the lectin had to be diluted and concentrated in the buffer containing the ligand. Crystals were obtained with the same procedure as described above.

Data collection and processing

Preliminary X-ray data collection to 2.6 Å resolution was carried out at AstraZeneca R&D Mölndal, on a rotating copper anode (Rigaku RU300 HB) equipped with a Mar 345 image plate (at $T \approx 100$ K). The crystals belonged to space group $P4_32_12$ with cell dimensions of $a = b = 81.8$ Å and $c = 126.1$ Å, containing one molecule per asymmetric unit. After the structure of the lactose complex had been solved, a high-resolution data set (including reflections to 1.45 Å resolution) was obtained at the synchrotron MAX-lab II in Lund, at beamline I711, also equipped with a Mar 345 image plate. Data were collected at cryogenic temperature (100 K). The wavelength for data collection was set to 1.03 Å and the crystal to detector distance to 200 mm. Data collection proceeded in two steps. First, a high-resolution data set was collected covering 90°, with oscillation ranges of 0.5°. A quick second data collection sweep covered another 90° of reciprocal space (oscillation range 1.25°), ensuring high completeness also of the low-resolution data.

The data for the ECL complex with fucosyllactose were collected in a similar way, however, with oscillation ranges of 1.0° and 1.25°, respectively. The crystal diffracted to 1.7 Å resolution.

All data were processed and scaled with XDS.[30–32] Data collection statistics are summarized in Table 1.

Structure determination and refinement

The structure of ECL in complex with lactose was solved by molecular replacement, using the program AMoRE.[41] Structure determination was straightforward, without the need to adjust many parameters. As a search model, we took the crystal structure of ECorL[7] (PDB entry 1AX0), from which the calcium and manganese ions, the carbohydrate ligand and all water molecules had been removed. Refinement was carried out with CNS.[42] After the first refinement cycle, the calcium and manganese ions were added to the model. Simulated-annealing omit maps were calculated in order to remove any model bias from the electron density maps used for model building. The electron density maps clearly showed the presence of the lactose ligand, even at the earliest stages of refinement. Refinement (in the first rounds including simulated annealing) altered with cycles of manual rebuilding using the computer graphics package O.[43] Parameter and topology files for the carbohydrates were taken from the HICUP server,[44] and some of the weights were adjusted as necessary. After several rounds of refinement, water molecules were added to the structure, if the corresponding ($F_{obs} - F_{calc}$) density was at least 3σ and geometric requirements for hydrogen bonding were fulfilled. In the final stages of the refinement, alternative conformations were defined for residues 9, 10, 12, 95, 159, 173, 180, 234 as well as for glucose O6′. The electron density for glucose O1′ was found to be consistent with a mixture of the α and β-anomers. As the *R*-factors increased noticeably for data beyond 1.58 Å resolution and the electron density maps at higher resolution did not reveal much additional information, the data were cut at 1.58 Å resolution in the final refinement cycle.

The structure of the ECL fucosyllactose complex was refined using the apo-structure of the ECL lactose complex as a basis for rigid body refinement, followed by simulated annealing. Refinement then proceeded as described above for the ECL lactose complex, however, without refining alternative conformations for protein side-chains. The fucosyllactose ligand was built into the density in two steps, first the lactose unit and later the fucose ring.

Structure comparisons

Comparisons of different lectin complexes were usually based on a superposition of the respective C^{α} coordinates of amino acid residues 85–220 using the program O.[43] For ECorL, the comparative analysis was based on PDB entry 1AX1 (complex with lactose at 1.95 Å resolution),[9] where not stated otherwise. Alternatively to this "standard" superposition of residues 85–220, also alignments involving only three conserved residues, e.g. Asp89, Gly107 and Asn133 or Asp89, Asn133 and Ala218 were tested, with essentially the same results. The comparison of the fucosyllactose complexes of ECL and UEA-II[25] (1QOT, monomer III) was based on a superposition of the galactose residues.

Figures

Figures 3–5(a) and (b) were generated with Bobscript (Figure 3)[45] and Molscript[46] (Figures 4 and 5), respectively. Figure 5(c) and (d) were generated with ISIS™/Draw 2.4.

Atomic coordinates

The refinement statistics are summarized in Table 1. The coordinates and structure factors are deposited with the Protein Data Bank (accession codes 1GZC and 1GZ9).

Disclaimer

Certain commercial materials, instruments, and equipment are identified in this manuscript in order to specify the experimental procedure as completely as possible. In no case does such identification imply a recommendation or endorsement by the National Institute of Standards and Technology nor does it imply that the materials, instruments, or equipment identified is necessarily the best available for the purpose.

Acknowledgments

We thank Niamh Roche for setting up the first successful crystallization experiments, Jonas Ångström for stimulating discussions and Pierre Thibault for making the first sequence of an *Erythrina cristagalli* lectin available to us. We are further grateful to Tomas Lundqvist for data collection time and assistance at AstraZeneca and to Yngve Cerenius for support at MAX-lab in Lund. Lise-Lotte Olsson was of great help when sorting out problems in the fucosyllactose parameter file. This study has been supported by grants

from the Glycoconjugates in Biological Systems program from the Swedish National Foundation for Strategic Research (C.S., U.K.), the Swedish Medical Research Council (Grants No. 12628; S.T.; 14113, A0692; C.L.N.), the Swedish Cancer Foundation, the IngaBritt and Arne Lundberg Foundation, and the Knut and Alice Wallenberg Foundation.

References

1. Sharon, N. & Lis, H. (1993). Carbohydrates in cell recognition. *Sci. Am.* 268, 82–89.
2. Science (2001). Review series. *Science*, 291, 2337–2378.
3. Sharon, N. & Lis, H. (2001). The structural basis for carbohydrate recognition by lectins. In *The Molecular Immunology of Complex Carbohydrates-2* (Albert, M., ed.), Academic/Plenum, New York.
4. Lis, H. & Sharon, N. (1998). Lectins: carbohydrate-specific proteins that mediate cellular recognition. *Chem. Rev.* 98, 637–674.
5. Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta*, 1383, 9–36.
6. Moreno, E., Teneberg, S., Adar, R., Sharon, N., Karlsson, K.-A. & Ångström, J. (1997). Redefinition of the carbohydrate specificity of *Erythrina corallodendron* lectin based on solid-phase binding assays and molecular modeling of native and recombinant forms obtained by site-directed mutagenesis. *Biochemistry*, 36, 4429–4437.
7. Shaanan, B., Lis, H. & Sharon, N. (1991). Structure of a legume lectin with an ordered N-linked carbohydrate in complex with lactose. *Science*, 254, 862–866.
8. Elgavish, S. & Shaanan, B. (1998). Structures of the *Erythrina corallodendron* lectin and of its complexes with mono- and disaccharides. *J. Mol. Biol.* 277, 917–932.
9. Lemieux, R. U., Ling, C.-C., Sharon, N. & Streicher, H. (2000). The epitope of the H-Type 2 trisaccharide recognized by *Erythrina corallodendron* lectin. Evidence for both attractive polar and strong hydrophobic interactions for complex formation involving a lectin. *Israel J. Chem.* 40, 167–176.
10. Surolia, A., Sharon, N. & Schwarz, F. P. (1996). Thermodynamics of monosaccharide and disaccharide binding to *Erythrina corallodendron* lectin. *J. Biol. Chem.* 271, 17697–17703.
11. Gupta, D., Cho, M., Cummings, R. D. & Brewer, C. F. (1996). Thermodynamics of carbohydrate binding to galectin-1 from Chinese hamster ovary cells and two mutants. A comparison with four galactose-specific plant lectins. *Biochemistry*, 35, 15236–15243.
12. Srinivas, V. R., Bhanuprakash Reddy, G. & Surolia, A. (1999). A predominantly hydrophobic recognition of H-antigenic sugars by winged bean acidic lectin: a thermodynamic study. *FEBS Letters*, 450, 181–185.
13. Mandal, D. K., Bhattacharyya, L., Koenig, S. H. & Brown, R. D., III (1994). Studies of the binding specificity of concanavalin A. Nature of the extended binding site for Asparagine-linked carbohydrates. *Biochemistry*, 33, 1157–1162.
14. Gupta, D., Dam, T. K., Oscarson, S. & Brewer, C. F. (1997). Thermodynamics of lectin–carbohydrate interactions. *J. Biol. Chem.* 272, 6388–6392.
15. Toone, E. J. (1994). Structure and energetics of protein–carbohydrate complexes. *Curr. Opin. Struct. Biol.* 4, 719–728.
16. Bradbrook, G. M., Forshaw, J. R. & Pérez, S. (2000). Structure/thermodynamics relationships of lectin–saccharide complexes. *Eur. J. Biochem.* 267, 4545–4555.
17. Iglesias, J. L., Lis, H. & Sharon, N. (1982). Purification and properties of a D-galactose/*N*-acetyl-D-galactosamine specific lectin from *Erythrina cristagalli*. *Eur. J. Biochem.* 123, 247–252.
18. Lis, H., Joubert, F. T. & Sharon, N. (1985). Isolation and properties of *N*-acetyllactosamine-specific lectins from nine Erythrina species. *Phytochemistry*, 24, 2803–2809.
19. Teneberg, S., Ångström, J., Jovall, P.-Å. & Karlsson, K.-A. (1994). Characterization of binding of Galβ4GlcNAc-specific lectins from *Erythrina cristagalli* and *Erythrina corallodendron* to glycosphingolipids. *J. Biol. Chem.* 269, 8554–8563.
20. Bhattacharyya, L., Haraldsson, M., Sharon, N., Lis, H. & Brewer, F. (1989). Binding and precipitating activities of *Erythrina* lectins with complex type carbohydrates and synthetic cluster glycosides. A comparative study of the lectins from *E. corallodendron*, *E. cristagalli*, *E. flabelliformis*, and *E. indica*. *Glycoconj. J.* 6, 141–150.
21. Mann, M., Højrup, P. & Roepstorff, P. (1993). Use of mass spectrometric molecular weight information to identify proteins in sequence databases. *Biol. Mass Spectrom.* 22, 338–345.
22. Yates, J. R., Speicher, S., Griffin, P. R. & Hunkapiller, T. (1993). Peptide mass maps: a highly informative approach to protein identification. *Anal. Biochem.* 214, 397–408.
23. James, P., Quadroni, M., Carafoli, E. & Gonnet, G. (1994). Protein identification in DNA databases by peptide mass fingerprinting. *Protein Sci.* 3, 1347–1350.
24. Mann, M. & Wilm, M. (1994). Error-tolerant identification of peptides in sequence databases by peptide sequence tags. *Anal. Chem.* 66, 4390–4399.
25. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* 26, 283–291.
26. Kabsch, W. & Sander, C. (1983). Program to predict secondary structure and solvent exposure of protein from atomic coordinates as given by the Brookhaven Protein Data Bank. *Biopolymers*, 22, 2577–2637.
27. Srinivas, V. R., Bhanuprakash Reddy, G., Ahmad, N., Swaminathan, C. P., Mitra, C. P. & Surolia, A. (2001). Legume lectin family, the natural mutants of the quaternary state, provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys. Acta*, 1527, 102–111.
28. Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998). Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å resolution. *J. Mol. Biol.* 276, 787–796.
29. Sharma, V., Srinivas, V. R. & Surolia, A. (1996). Cloning and sequencing of winged bean (*Psophocarpus tetragonolobus*) basic agglutinin (WBA I): presence of second glycosylation site and its implications in quaternary structure. *FEBS Letters*, 389, 289–292.

30. Sharma, V. & Surolia, A. (1997). Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *J. Mol. Biol.* **267**, 433–445.
31. Loris, R., Stas, P. P. G. & Wyns, L. (1994). Conserved water in legume lectin crystal structures. *J. Biol. Chem.* **269**, 26722–26733.
32. Lesk, A. M. C. & Chothia, C. (1980). How different amino acid sequences determine similar protein structures: the structure and evolutionary dynamics of the globins. *J. Mol. Biol.* **136**, 225–270.
33. Weis, W. I. & Drickamer, K. (1996). Structural basis of lectin–carbohydrate recognition. *Annu. Rev. Biochem.* **65**, 441–473.
34. Rini, J. M. (1995). Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.
35. Loris, R., De Greve, H., Dao-Thi, M. H., Messens, J., Imberty, A. & Wyns, L. (2000). Structural basis of carbohydrate recognition by lectin II from *Ulex europaeus*, a protein with a promiscuous carbohydrate-binding site. *J. Mol. Biol.* **301**, 987–1002.
36. Pereira, M. E. A., Gruezo, F. & Kabat, E. A. (1979). Purification and characterization of lectin II from *Ulex europaeus* seeds and an immunochemical study of its combining site. *Arch. Biochem. Biophys.* **194**, 511–525.
37. Karlsson, K.-A. (1987). Preparation of total non-acid glycolipids for overlay analysis of receptors for bacteria and viruses and for other studies. *Methods Enzymol.* **138**, 212–220.
38. Kabsch, W. (1988). Automatic indexing of rotation diffraction patterns. *J. Appl. Crystallog.* **21**, 67–71.
39. Kabsch, W. (1988). Evaluation of single-crystal X-ray diffraction data from a position-sensitive detector. *J. Appl. Crystallog.* **21**, 916–924.
40. Kabsch, W. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Crystallog.* **26**, 795–800.
41. Navaza, J. (1994). AMoRE: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
42. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
43. Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
44. Kleywegt, G. J. & Jones, T. A. (1998). Databases in protein crystallography. *Acta Crystallog. sect. D*, **54**, 1119–1131.
45. Esnouf, R. M. (1997). An extensively modified version of MolScript that includes greatly enhanced coloring capabilities. *J. Mol. Graph. Model.* **15**, 132–134.
46. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.

Edited by W. Baumeister

(Received 13 February 2002; received in revised form 24 May 2002; accepted 29 May 2002)

doi:10.1016/j.jmb.2003.11.043 J. Mol. Biol. (2004) 335, 1227–1240

JMB

Available online at www.sciencedirect.com

SCIENCE DIRECT®

ELSEVIER

Structural Basis of Oligomannose Recognition by the *Pterocarpus angolensis* Seed Lectin

Remy Loris[1]*, Ivo Van Walle[1], Henri De Greve[1], Sonia Beeckmans[2] Francine Deboeck[3], Lode Wyns[1] and Julie Bouckaert[1]

[1]*Laboratorium voor Ultrastructuur, Instituut voor Moleculaire Biologie Building E, Vrije Universiteit Brussel and Vlaams Instituut voor Biotechnologie Pleinlaan 2, B-1050 Brussel Belgium*

[2]*Laboratorium voor Scheikunde der Proteïnen, Instituut voor Moleculaire Biologie Vrije Universiteit Brussel Pleinlaan 2, B-1050 Brussel Belgium*

[3]*Laboratorium voor Genetische Virologie, Instituut voor Moleculaire Biologie Vrije Universiteit Brussel Pleinlaan 2, B-1050 Brussel Belgium*

**Corresponding author*

The crystal structure of a Man/Glc-specific lectin from the seeds of the bloodwood tree (*Pterocarpus angolensis*), a leguminous plant from central Africa, has been determined in complex with mannose and five manno-oligosaccharides. The lectin contains a classical mannose-specificity loop, but its metal-binding loop resembles that of lectins of unrelated specificity from *Ulex europaeus* and *Maackia amurensis*. As a consequence, the interactions with mannose in the primary binding site are conserved, but details of carbohydrate-binding outside the primary binding site differ from those seen in the equivalent carbohydrate complexes of concanavalin A. These observations explain the differences in their respective fine specificity profiles for oligomannoses. While Man(α1-3)Man and Man(α1-3)[Man(α1-6)]Man bind to PAL in low-energy conformations identical with that of ConA, Man(α1-6)Man is required to adopt a different conformation. Man(α1-2)Man can bind only in a single binding mode, in sharp contrast to ConA, which creates a higher affinity for this disaccharide by allowing two binding modes.

Keywords: lectin; carbohydrate recognition; legume lectin; mannose

Introduction

Protein–carbohydrate recognition is a major form of inter-cellular communication and plays a role in many biologically important processes such as viral, bacterial, mycoplasmal and parasitic infections, targeting of cells and soluble components, fertilisation, cancer metastasis and growth and differentiation.[1]

The specific recognition of an (oligo)saccharide by a protein is a much more complex problem than other biologically relevant recognition processes such as protein–protein or protein–DNA interactions. The monosaccharide building blocks of a glycan are difficult to distinguish from each other due to the limited repertoire of functional groups involved. Apart from the occasional *N*-acetyl group or rarely a carboxylate group, one finds invariantly a large abundance of hydroxyl groups interspersed with small aliphatic patches. In addition, the glycosidic bonds between two monosaccharides are rather flexible, especially the 1-6 linkage. As a consequence, a high entropic cost limits the binding affinities that can be obtained. The combination of flexibility and the difficulty in distinguishing monomeric building blocks allows oligosaccharides to mimic each other structurally, making the task of specific recognition a truly difficult one.

The lectins from leguminous plants have been considered as a model system for studying the molecular basis of protein–carbohydrate interactions for several decades. Among all known

Abbreviations used: ConA, concanavalin A from *Canavalia ensiformis*; PAL, *Pterocarpus angolensis* Man/Glc-specific seed lectin; MAL, *Maackia amurensis* leuckoagglutinin; GS-IV, *Griffonia simplicifolia* lectin IV; UEA-II, *Ulex europaeus* lectin II; FRIL, Flt3 receptor interacting lectin from *Dolichos lablab*; LOL, *Lathyrus ochrus* lectin.

E-mail address of the corresponding author: reloris@vub.ac.be

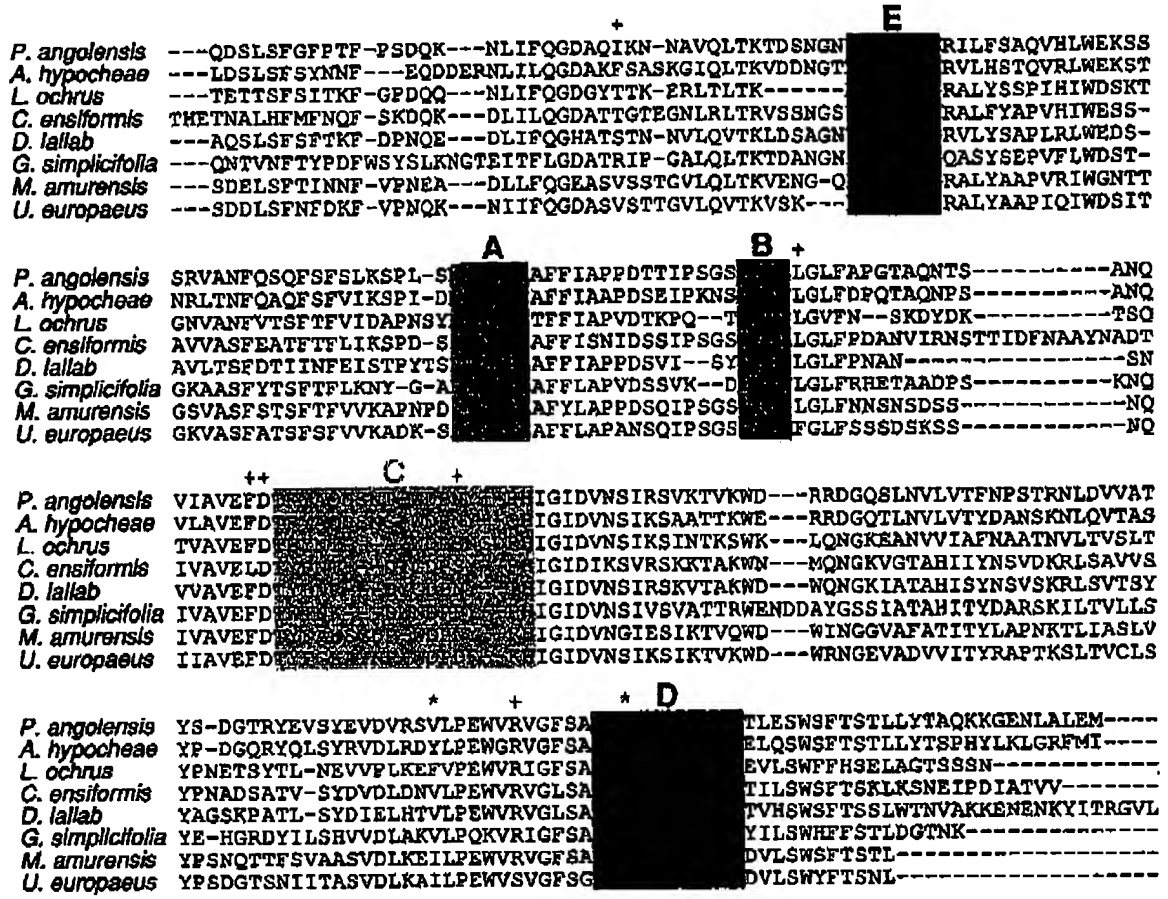


Figure 1. Amino acid sequence of the *P. angolensis* lectin. The amino acid sequence of PAL as used in the crystal structure determination is aligned with those of other legume lectins discussed in the text. Residues where the crystal structure conflicts with the cDNA-derived sequences are indicated by an asterisk (*). Positions where sequence variation was observed among the different cDNA clones sequenced are indicated by a +. The five stretches A–E that constitute the carbohydrate-binding site are indicated and shaded in different colours.

lectin families, animal, plant or microbial, they cover the widest range of carbohydrate specificities. Their carbohydrate recognition site consists of several loops, with differing degrees of variability.[2,3] The conformations of these loops are determined by the presence of a structural calcium ion and a transition metal ion,[4–6] the absence of which results in local unfolding and loss of carbohydrate-binding capacity. The vast amount of structural and biochemical data already available allows us to rationalise the structural determinants of carbohydrate specificity. Indeed, the crystal structures of more than 25 legume lectins have been determined[3,7–20] and are present in the Lectin Database†. Here, we present the crystal structure of the seed lectin (PAL) from the bloodwood tree (*Pterocarpus angolensis;* a leguminous plant from central Africa) in complex with a series of oligomannose ligands. This lectin belongs to the Man/Glc specificity group and has a preference for Man(α1-2)Man and Man(α1-3)[Man(α1-6)]Man (L. Buts & S.B., unpublished results). Its fine specificity differs from that of other known Man/Glc-specific lectins, while its amino acid sequence suggests some peculiarities in the carbohydrate-binding sites not seen in other legume lectins with related specificities.

† http://www.cermav.cnrs.fr/lectines/

Results and Discussion

Amino acid sequence and overall structure of PAL

The amino acid sequence was deduced from 14 cDNA clones (data not shown). The lectin gene starts with an ATG start codon followed by a signal sequence corresponding to 34 amino acid residues. This signal peptide ends with a putative signal peptide cleavage site between residues Ser and Gln.[21] This glutamine is the N-terminal residue

observed in all X-ray structures and is present in the form of a cyclic glutamine residue. This is in agreement with the fact that the N terminus of the affinity-purified protein was proven to be blocked (data not shown).

At several places, heterogeneity was observed in the cDNA sequences: Ile/Thr26, Leu/Pro108, Phe/Leu129, Asp/His130, Asn/Asp143 and Gly/Arg212. In the electron density maps of our crystal structures, these residues were identified as Ile26, Leu108, Phe129, Asp130, Asn143 and Arg212. Since the sequences were determined starting from mRNA, it can be assumed that all the corresponding protein variants are indeed synthesised. With the exception of Ile/Thr26, all these residues are located at or near the carbohydrate-binding site. Especially Phe/Leu129 and Asp/His130 are known to be critical for carbohydrate and metal-binding and are conserved in other legume lectins. It is possible, therefore, that some of the cDNA clones correspond to proteins that are closely related to PAL but are inactive or have altered carbohydrate specificity.

Figure 1 aligns the sequence of mature PAL as used in the crystal structure determination with those of other legume lectins that are discussed here. At two positions, the electron density was in disagreement with the cDNA-derived sequences: position 206 was interpreted as Val rather than Ala, since clear side-chain density was present. Val was chosen over Thr, as this side-chain is pointing towards a hydrophobic area without any potential hydrogen bond donors or acceptors. At position 220, Gly was modelled instead of Arg. There is no electron density for an Arg side-chain (including its C^{β} atom) in this otherwise well-defined region of the electron density map, and the presence of even a small side-chain at this position would clash with any sugar bound in the carbohydrate recognition site.

PAL has the highest level of sequence identity (64%) with the Man/Glc-specific lectin from *Arachis hypogae* (peanut), for which no crystal structure is available. Among those lectins for which the crystal structure is determined, the sialyllactose-specific MAL from *Maackia amurensis* (the Amur Maackia tree from Asia) is most closely related (43% sequence identity), while other Man/Glc-specific lectins such as ConA, FRIL and LOL show only 40–42% identity.

The overall structure of PAL has been described and is shown in Figure 2.[22] It consists of the typical legume lectin β-sandwich, the details of which are well known.[23]

All structures presented show good electron density for residues 1–238. Tyr239 was fit into the density, but the temperature factors for this residue remain high. Weak density is seen for residues Thr240 and Ala241. The bound carbohydrate molecules display clear electron densities in each of the complexes (Figure 3).

There are two legume lectin monomers in the

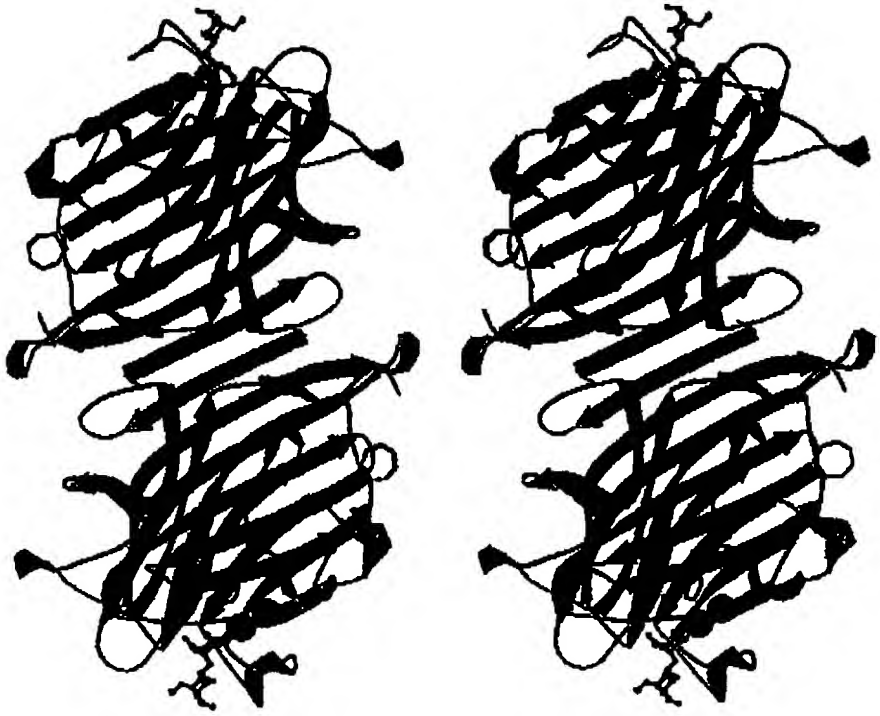


Figure 2. Overall structure of the *P. angolensis* lectin. Stereo cartoon representation of the PAL dimer (Man(α1-3)Man complex). One monomer is coloured orange, the other one yellow. Manganese ions are shown as light blue spheres and calcium ions as green spheres. Two bound molecules of Man(α1-3)Man are shown in ball-and-stick representation.

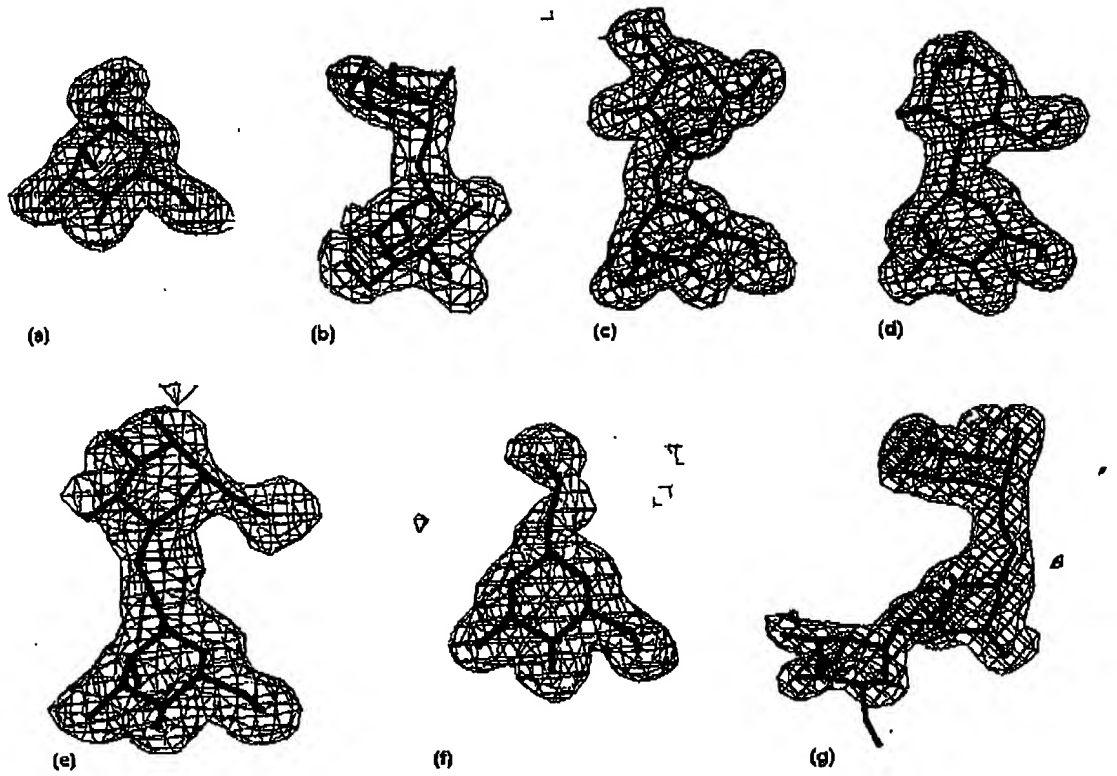


Figure 3. Experimental electron densities of the bound carbohydrates. (a) α-Me-Mannose, (b) Man(α1-2)Man, (c) Man(α1-3)Man, (d) Man(α1-4)Man, (e) Man(α1-6)Man, (f) Man(α1-6)[Man(α1-3)]Man in monomer A and (g) Man(α1-6)[Man(α1-3)]Man in monomer B. In each case, except (g), density in the binding site of the A monomer is shown. In all cases, the electron density map is an $F_{obs} - F_{calc}$ map contoured at 3.0σ and obtained by taking the final refined model and deleting the carbohydrates in both binding sites.

asymmetric unit, together forming the canonical dimer (Figure 2), as has been observed in many other dimeric and tetrameric legume lectins.[24,25] As a consequence of the crystal packing, the carbohydrate-binding sites of subunits A and B in the asymmetric unit of the crystals are not equivalent. The binding site of subunit A is involved in crystal packing. In all carbohydrate complexes, the ligand bound to the binding site of subunit A makes contacts with protein atoms from a symmetry mate. The binding site of subunit B on the other hand is not involved in crystal packing. In all but one of the carbohydrate complexes, binding seems to be identical with both subunits. The only exception is the Man(α1-3)[Man(α1-6)]Man complex, where the conformation adopted in binding site B is prevented in site A due to a steric hindrance from a symmetry-related protein molecule.

Carbohydrate-binding site

The carbohydrate-binding site of PAL can be described as one with a primary binding site that recognises glucose and mannose. This primary binding site lies in the centre of a shallow groove on the surface of the protein (Figure 4(a)). Extensions to the sugar bound in the primary binding site (M) can be made in two directions, following O1 or O2. Independent of the exact linkage that is present, these additional sugar residues occupy two particular regions of the lectin, which we call the −1 (attached to O2) and +1 (attached to O1) subsites. In the case of longer oligosaccharides, we speak of additional +2, +3; ... or −2, −3, ... subsites depending on their position relative to the primary binding site.

The carbohydrate-binding site of all legume lectins consists of residues belonging to five polypeptide stretches (termed A to E according to Sharma & Surolia[2]) (Figure 4(b)), which vary to different degrees between lectins with different specificities. Stretches A and B contain an essential aspartate residue (invariantly preceded by a *cis*-peptide bond) and backbone NH group (usually from a glycine residue; Gly106 in PAL), respectively. The conformations of these two stretches do not vary much among different lectins. In the current structures, this picture is confirmed.

Stretch C is the metal-binding loop and wraps around the structurally important calcium and manganese ions. In the known crystal structures, five different loop sizes (12–16 residues) adopting

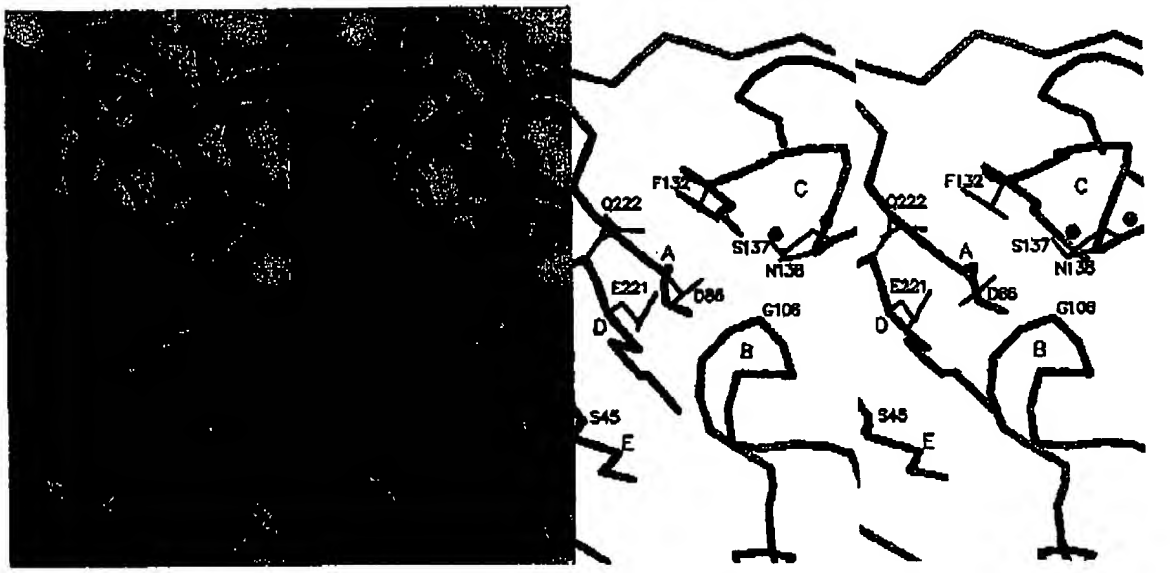


Figure 4. Architecture of the carbohydrate-binding site. (left) Stereo view of a CPK model of PAL with the four stretches that constitute the carbohydrate-binding site coloured: blue for stretch A containing *cis*-Asp86, orange for loop B containing Gly106, yellow for metal-binding loop C and green for specificity loop D. These colours are similar to those used in Figure 1 and are maintained in all the other Figures. Superimposed is the ball-and-stick model of a virtual tetrasaccharide Man(α1-2)Man(α1-6)[Man(α1-3)]Man indicating the different subsites (M, primary binding site and −1, +1 and +2 downstream and upstream subsites. (right) Stereo MOLSCRIPT representation showing the five stretches A–E that make up the carbohydrate-binding site. Side-chains of residues important for carbohydrate-binding are shown in ball-and stick and are labelled. The calcium and manganese ions are shown as large green and yellow spheres, respectively.

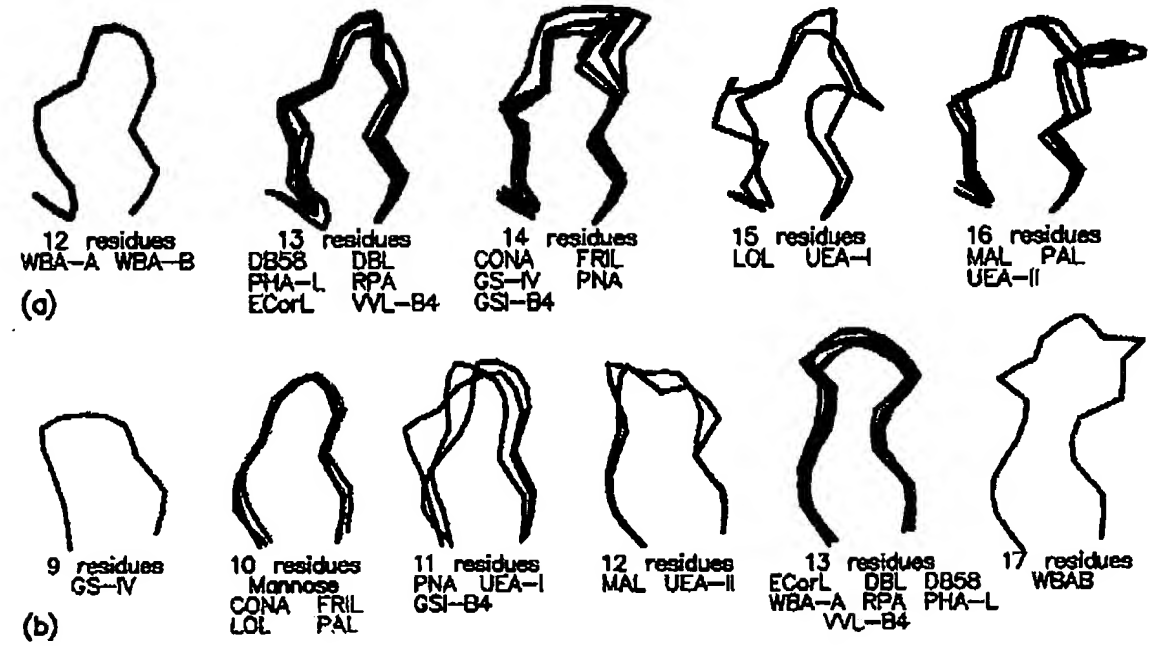


Figure 5. Loop conformations in the carbohydrate-binding sites of legume lectins. (a) conformations observed for loop C and (b) conformations observed in loop D suggesting the use of canonical loop conformations to modulate carbohydrate specificity in a way similar to that seen in the CDR loops of antibodies.

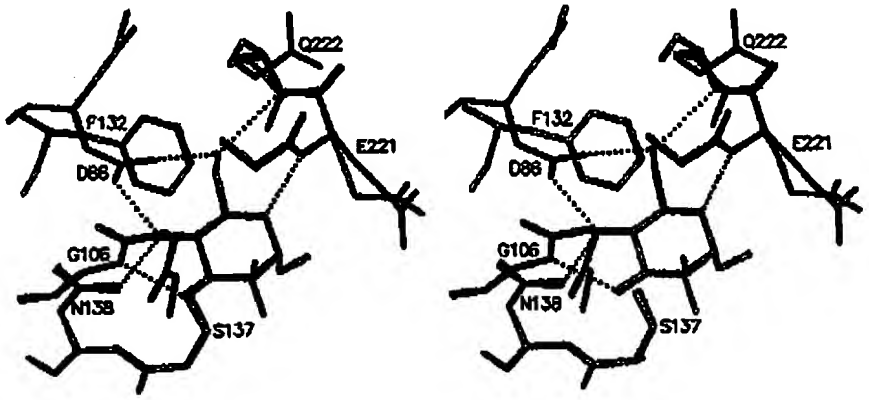


Figure 6. Interaction of mannose in the primary binding site. Stereo view of the superposition of the monosaccharide-binding sites of PAL complexes with MeαMan coloured and MeαGlc light grey. Selected residues are labelled. The complexes are structurally identical with the exception of the orientation of the O2 atom of the sugars. The difference in conformation for Glu221 is most likely due to the less well defined electron density of this residue in the MeαG structure and is probably not of biological relevance.

a total of eight different conformations are observed (Figure 5(a)). There is no simple relationship between monosaccharide specificity and the length and conformation of loop C. In PAL, this loop has a length of 16 residues and is identical with that seen in the two sialyllactose-specific lectins from *M. amurensis*[18] and in the chitobiose-specific lectin II from common gorse (*Ulex europaeus*).[19] The other Man/Glc-specific lectins for which crystal structures are available have loop lengths of either 14 (FRIL, ConA and related lectins) or 15 (LOL and related lectins) residues. While the backbone conformation of loop C is not a determinant for monosaccharide specificity, specific side-chains on this loop nevertheless do influence the nature of the sugar that can be accommodated in the binding site (see below).

In contrast to stretches A–C, stretch D does not interact directly with the structural calcium ion. It is highly variable in length, conformation and sequence and is often referred to as the monosaccharide specificity loop.[3] It is thought to be the prime determinant for monosaccharide as well as oligosaccharide specificity. In the PAL structure, the conformation adopted by this loop of ten residues is identical with that found in all other known crystal structures of Man/Glc-specific lectins[15–17,26–28] (Figure 5(b)), supporting this notion.

Finally, stretch E is found to interact with a bound carbohydrate in only a few cases, where it is part of the −1 subsite: *M. amurensis* leuckoagglutinin (MAL) in complex with sialyllactose,[19] *Griffonia simplicifolia* lectin IV (GS-IV) in complex with the Le^b tetrasaccharide[29] and ConA in complex with Man(α1-2)Man.[30] In most complexes of PAL, it is not involved in carbohydrate-binding, the exception being the Man(α1-2)Man complex.

Interactions in the primary binding site

Methyl-α-D-mannopyrannoside (MeαMan) binds to the primary binding site in the same way as has been observed repeatedly for other Man/Glc-specific lectins (Figure 6).[15,20,26–28] The sugar interacts with the protein *via* a series of hydrogen bonds with the conserved Asp/Gly/Asn triad as well as with the backbone of specificity loop D (hydrogen bond between Glu221 and O5). In addition, the side-chain of Phe132 from the metal-binding loop C stacks favourably upon the sugar ring, while Gly220 and the side-chain of Glu221 also make favourable van der Waals contacts with the sugar.

As observed for most other Glc/Man-specific legume lectins, PAL has twice the affinity for MeαMan (our unpublished results) than for MeαGlc. In our crystal structure, the axial O2 of mannose makes van der Waals contacts with the C^α atoms of Gly106 and Gly220, possibly forming CH···O hydrogen bonds (C–O distances 3.7 Å and 3.5 Å, respectively). This interaction is absent from the MeαGlc complex[22] due to the equatorial orientation of O2 in this sugar. As a consequence, a small void is present that is not filled with an ordered water molecule.

Non-specific subsite interactions: Man(α1-3)-Man, Man(α1-4)Man and Man(α1-6)Man

The affinities of PAL for Man(α1-3)Man, Man-(α1-4)Man and Man(α1-6)Man are essentially identical with that for MeαMan (our unpublished results). Clear density for the disaccharides has been observed in both subunits. The conformation adopted by the disaccharide is in each case identical for both subunits, despite the involvement of the binding site of subunit A in crystal packing. All bind with their non-reducing mannose in the

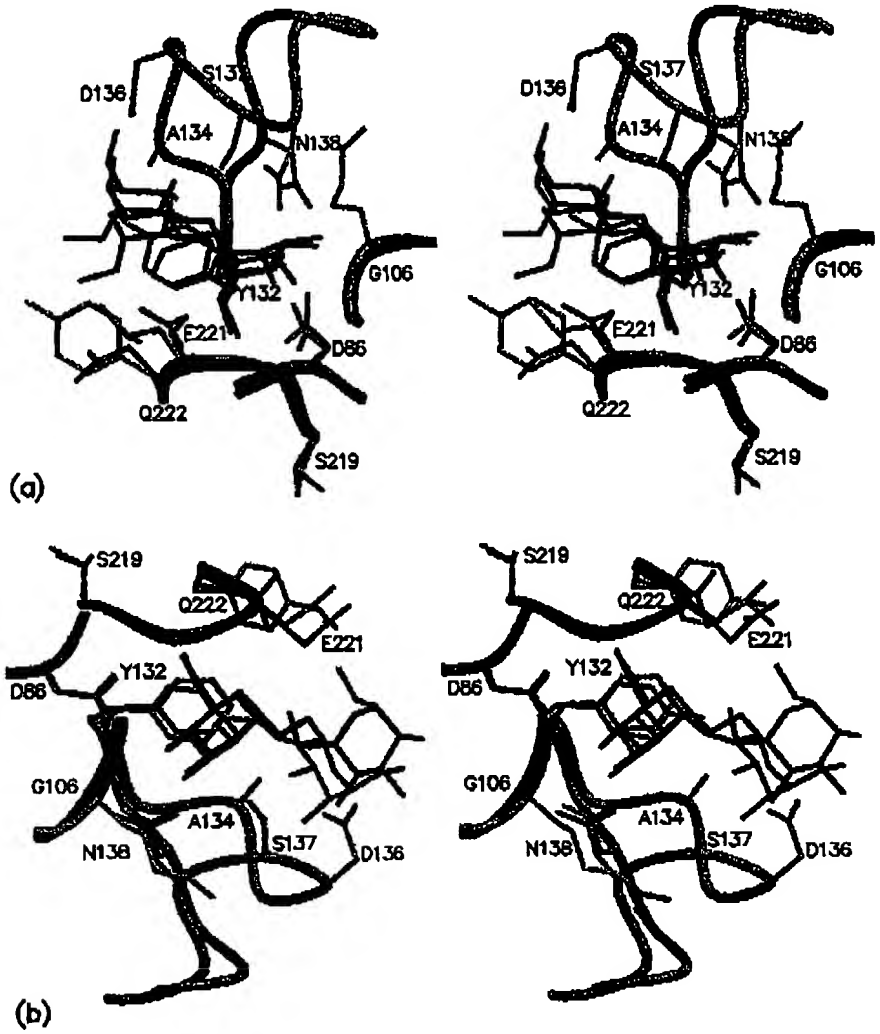


Figure 7. Recognition of Man(α1-3)Man and Man(α1-6)Man by PAL. (a) Stereo view of the PAL-binding site in complex with Man(α1-3)Man. The disaccharide is shown in blue and loops of PAL are coloured as in Figure 2. The equivalent loops and side-chains of the ConA-Man(α1-3)Man complex are shown in light grey. The ConA-bound disaccharide is shown in black. (b) A similar stereo view of the PAL-Man(α1-6)Man complex superimposed on the ConA-Man(α1-6)Man complex.

primary binding site, thus occupying subsite +1. The interactions with this subsite are sufficiently favourable to allow each of the disaccharides to adopt a unique conformation, but insufficient to result in an enhanced affinity. All three disaccharides make at least one additional hydrogen bond with amino acid residues in the +1 site. This situation is similar to what has been observed for ConA,[31] where Man(α1-6)Man binds with the same affinity as MeαMan, despite an additional hydrogen bond from the protein to the reducing mannose.

Figure 7 compares the binding modes of Man(α1-3)Man and Man(α1-6)Man with those observed previously in their complexes with ConA.[31] Man(α1-3)Man binds in an identical conformation and orientation to ConA and PAL, although the residues that make up subsite +1 differ in both proteins (Figure 7(a)). Man(α1-6)Man recognition on the other hand is different in both proteins, and the disaccharides adopt different low-energy conformations (Figure 7(b)). The conformation found in ConA cannot be adopted in the PAL structure due to clashes with the metal-binding loop C, notably the side-chains of Ala134 and Asp136. The conformation found in PAL could, on the other hand, fit in the binding site of ConA, provided that the α-methyl group on O1 would be reoriented, but no hydrogen bond would be formed in the +1 site.

Specificity for α *versus* β linkages

Similar to other Man/Glc-specific legume lectins,

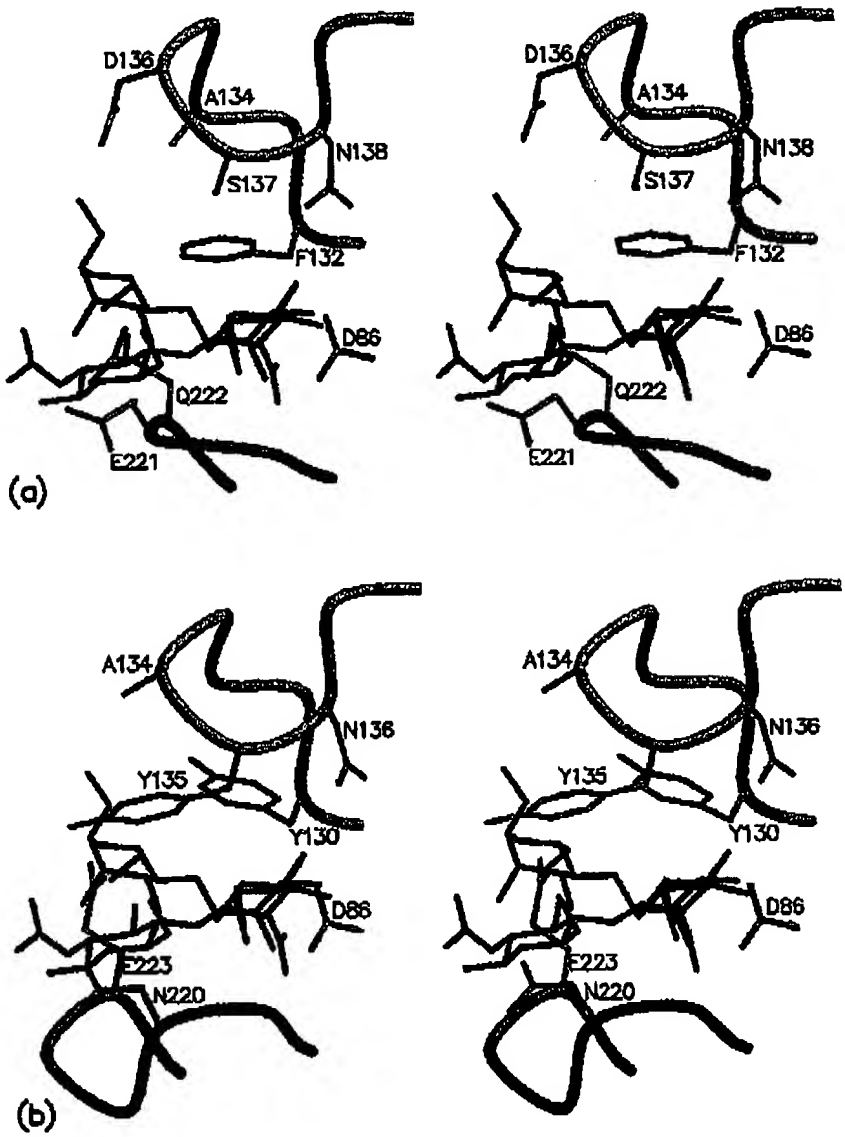


Figure 8. Binding of Man(α1-4)Man by PAL. (a) A stereo view of the PAL-binding site in complex with Man(α1-4) Man (blue). Loops are coloured as in Figure 2. In black, the co-ordinates of GlcNAc(β1-4)GlcNAc as seen in the binding site of UEA-II are superimposed. In PAL, the β anomeric configuration is made impossible due to residues Glu221 and Gln222 in the specificity loop of PAL. (b) A stereo view of the UEA-II-binding site in complex with GlcNAc(β1-4) GlcNAc (blue). In black, the co-ordinates of Man(α1-4)Man as seen in the binding site of PAL are superimposed. In UEA-II, the α anomeric configuration is made impossible due to the bulky side-chain of Tyr135, which is replaced by Ser137 in PAL.

the *P. angolensis* lectin has a strict requirement for α-linkages. The chitobiose-specific lectin II from *U. europaeus* (UEA-II), on the other hand, can accommodate only a β-linkage. From the different crystal structures of UEA-II-carbohydrate complexes,[19] it has been learned that the *N*-acetyl groups of chitobiose are not crucial for binding and that its primary binding site can accommodate mannose and glucose as efficiently.

Figure 8 compares the binding of Man(α1-4)Man on PAL to that of GlcNAc(β1-4)GlcNAc (chitobiose) to UEA-II. The bulky Tyr135 in metal-binding loop C prevents formation of an α-linkage in the binding site of UEA-II and is replaced by the small Ser137 in PAL. Otherwise, the backbone conformations of this loop are very similar in PAL and UEA-II. On the other hand, Glu221 in specificity loop D of PAL sterically prevents a β-linkage from being accommodated in the binding site of this protein. In order to allow for a β-linkage, Glu221 needs to be at least truncated to Gly. In UEA-II, the specificity loop is two residues longer and

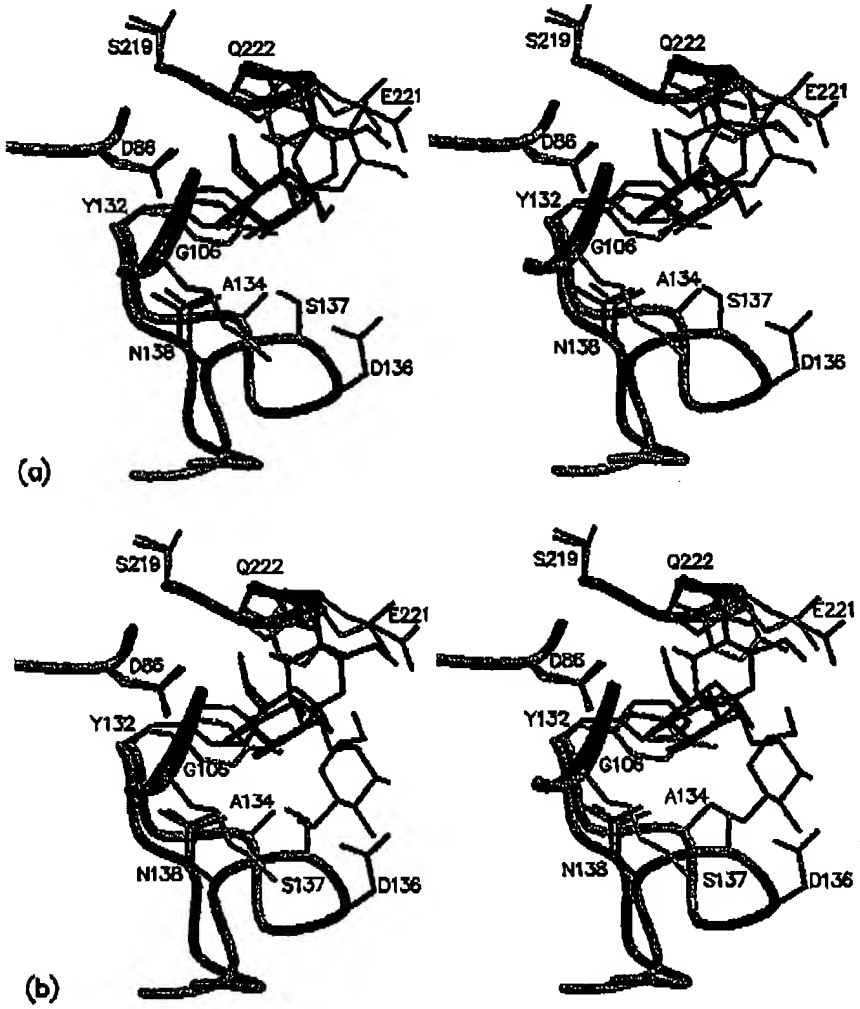


Figure 9. Recognition of Man(α1-2)Man by PAL. (a) A stereo view of the PAL-Man(α1-2)Man complex superimposed on the equivalent ConA complex (non-reducing mannose bound in subsite −1). Colouring is as in Figure 3. (b) An identical view of PAL, but superimposed on the ConA-Man(α1-2)Man complex with the second mannose in subsite +1.

locally adopts a different conformation, allowing for a β-linked disaccharide to be bound. Thus, the selection for α or β-glycosidic linkages appears to be controlled by a few specific amino acid residues and does not depend critically on the type of metal-binding loop, as was suggested earlier.[19]

Man(α1-2)Man binds with its reducing mannose in the primary binding site

In a typical Man/Glc-specific legume lectin, O2 of a bound MeαMan points towards the solvent. This means that, at least in theory, Man(α1-2)Man would be able to bind with both its reducing or its non-reducing mannose in the primary binding site. Such a situation is observed in the complex of concanavalin A (ConA) with Man(α1-2)Man,[30] and has been argued to contribute significantly to the enhanced affinity of this lectin for Man(α1-2)Man compared to MeαMan.[32]

PAL has a 2.5-fold higher affinity for Man(α1-2)-Man compared to MeαMan (our unpublished results). Nevertheless, in the current Man(α1-2)-Man complex only a single binding mode is observed, with the reducing mannose molecule in the primary binding site and the second mannose molecule in subsite −1. The second binding mode in the +1 subsite is not possible in PAL, as it would lead to a steric clash with the two residues longer metal-binding loop C, in particular with the side-chains of Ala134, Asp136 and Ser137 (Figure 9).

In the observed binding mode to PAL, the conformation of the disaccharide roughly resembles the equivalent one found in the ConA complex, but is not identical. The difference in conformation

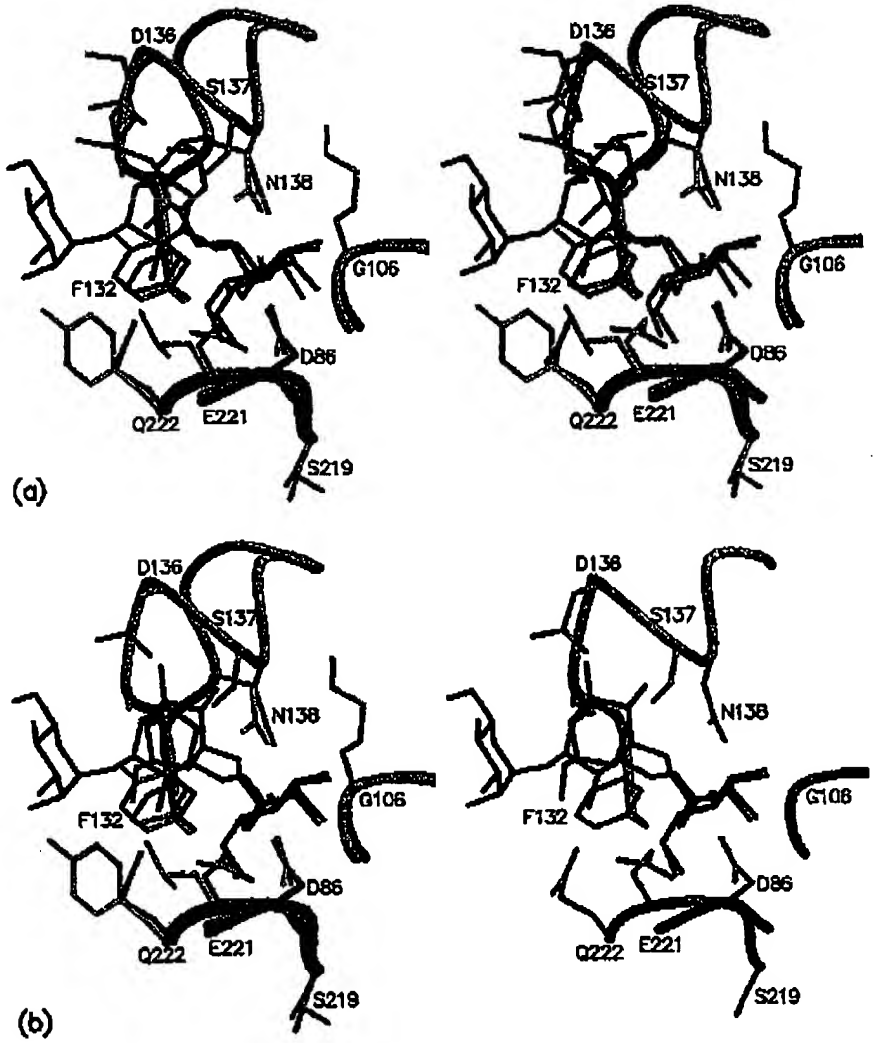


Figure 10. Recognition of the trisaccharide Man(α1-6)[Man(α1-3)]Man by PAL. Colouring is as in Figure 3. (a) Man(α1-6)[Man(α1-3)]Man bound by PAL. In light grey, the ConA-Man(α1-6)[Man(α1-3)]Man complex is superimposed (the sugar itself in black). (b) An identical view of the Man(α1-6)[Man(α1-3)]Man-PAL complex, but with Man(α1-6)Man superimposed in black.

is due to a single amino acid substitution in loop B: Gly104 in PAL is replaced by Thr226 in ConA, forcing a 30° rotation around Phi (defined as the torsion angle O5–C1–O1–C'2) to prevent a van der Waals clash and to establish a hydrogen bond between O2 of Man(−1) and Thr226(OG1) (Figure 9).

Context-dependent protein–carbohydrate interactions in the Man(α1-3)[Man(α1-6)]-Man complex

Man(α1-3)[Man(α1-6)]Man is the only carbohydrate for which there is a significant difference in binding in the two crystallographically independent lectin monomers. In the binding site of monomer B, which is not involved in crystal packing, clear density is seen for the whole trisaccharide. Outside the primary binding site, five direct and one water-mediated hydrogen bonds are made between the protein and the carbohydrate.

In the binding site of monomer A, on the other hand, electron density is observed only for the mannose in the primary binding site. As there is no trace of density corresponding to the original Man(α1-3)Man (that was present before the soak with the trimannose), it has to be assumed that the trimannose is indeed bound in the binding site. The obviously favourable conformation of the trisaccharide as observed bound to monomer B is not accessible in monomer A because of severe steric conflicts with a symmetry-related protein

Table 1. X-ray data collection and refinement statistics

Carbohydrate	MeαMan	Man(α1-2)Man	Man(α1-3)Man	Man(α1-4)Man	Man(α1-6)Man	Man(α1-3)[Man(α1-6)]Man
Unit cell parameters						
a (Å)	56.82	56.32	56.88	57.28	56.88	56.85
b (Å)	83.67	82.98	83.03	83.37	83.38	83.61
c (Å)	123.00	122.00	122.94	122.96	122.57	122.96
Beamline	BW7A	X31	BW7B	X31	X11	BW7A
Resolution(Å)	1.70	2.20	1.75	2.05	2.05	1.85
N_{meas}	210,120	101,201	395,205	209,488	160,081	232,267
N_{unique}	64,245	25,266	57,268	36,449	37,147	50,664
Completeness (%)	97.6	84.9	98.6	99.9	97.3	99.7
R_{merge}	0.051	0.102	0.177	0.119	0.111	0.081
$< I/\sigma(I) >$	19.20	8.88	13.05	11.67	9.79	17.00
R	0.1889	0.1776	0.1837	0.1841	0.1824	0.1805
R_{free}	0.2154	0.2178	0.2064	0.2179	0.2161	0.2052
PDB code	1ukg	1q8o	1q8p	1q8q	1q8 s	1q8v

molecule. Apparently, when specific subsite interactions cannot be made, the trisaccharide is bound in a disordered way, similar to that observed in the complex of pea lectin with the same trisaccharide.[28]

Concanavalin A has a 40 times higher affinity for the trisaccharide than for MeαMan.[33,34] PAL favours the trisaccharide by only a factor of four (our unpublished results). Figure 10 compares trimannose-binding by ConA[35,36] and monomer B of PAL. Although there are many differences for the amino acid residues outside the primary binding site, some of the global features of the binding of the trisaccharide are remarkably similar for both lectins. In both cases, the non-reducing mannose that is linked to O6 is found in the primary binding site. The "middle" (reducing) mannose makes extensive interactions with the protein. These interactions require, however, the presence of the third mannose molecule (non-reducing O3-linked). The details of the binding, such as specific hydrogen bonds formed outside the primary binding site, nevertheless differ markedly for ConA and PAL.

Materials and Methods

Purified *P. angolensis* seed lectin (PAL) was available from a previous study.[22] Internal amino acid sequences were obtained from four peptides after digestion of the purified *P. angolensis* lectin with trypsin. The amino acid sequences of these peptides (SPLSNGADGIAFFIA, and WDPDYPH) were used to design degenerate primers. Total RNA was extracted from ripening *P. angolensis* seeds (collected in Zimbabwe and stored at −80 °C) using the RNeasy Plant Mini Kit (Qiagen) and converted into double-stranded cDNA using the 5′RACE System for rapid amplification of cDNA ends (GibcoBRL) and the Q_T primer.[37] This cDNA was amplified with the degenerate forward primer Muk-2:

5′-GGIGCIGAYGGIATHGCITTYTT-3′

and the reverse primer Muk-5:

3′-ACCCTRGGICTRATRGGICT-5′

After amplification, the PCR fragments were cloned into the pUC18 using the Sure-Clone-Ligation kit (Amersham Pharmacia Biotech), and sequenced.

Specific primers Muk-6:

5′-CATCGCACCGCCGGATACTAC-3′

Muk-8:

3′-GAGTTTTGTGCAGGCGCTTGG-5′

and Muk-9:

3′-GATCCCGAGAAACGTGGACC-5′

were designed on the basis of the sequence information gathered from the PCR fragment obtained with the degenerate primers Muk-2 and Muk-5, to amplify and clone the nucleotide sequence corresponding to the 3′-end and the 5′-end of the *P. angolensis* mRNA.

The 3′-end of the cDNA was amplified with the Muk-6 and the Q_0 primer.[37] The 5′-end of the *P. angolensis* mRNA was amplified after adding a poly(C) tail to the 5′-end of the cDNAs (GibcoBRL). An initial PCR amplification was carried out with Muk-8 and RAAP (GibcoBRL), followed by a second PCR amplification using the nested primers Muk-9 and AUAP (GibcoBRL). The PCR fragments were cloned into the pUC18 using the Sure-Clone-Ligation kit (Amersham Pharmacia Biotech) and sequenced.

Full length cDNAs were subsequently amplified using primers Muk-32:

5′-CCCAAATATAATAAAAAGCGCTACCCA-TCTC-3′

(located in the 5′-untranslated sequence) Muk-14:

5′-CCTCCCTCTCCTTGCCTTCTTCA-3′:

(located in the signal peptide) or Muk-15:

5′-ATGCTACTGAACAAAGCATACT-3′

(located in the signal peptide) in combination with the Q_0 primer.[37] The PCR fragments were cloned in pUC18 and sequenced.

Sequencing was performed following the dideoxynucleotide termination method of Sanger *et al.*[38] using the Thermo Sequenase Radiolabeled Termination Cycle Sequencing kit (Amersham Pharmacia Biotech). Separation of the resulting DNA fragments was performed

on a 5% (w/v) polyacrylamide gel containing 8 M urea, and Tris-borate buffer (100 mM Tris, 100 mM boric acid). Detection of the sequencing ladders was achieved by autoradiography.

Crystallisation and data collection

All carbohydrates were purchased from Dextra Laboratories (Reading, UK) and are methylated on their reducing O1, even when not specifically mentioned otherwise. Crystals of PAL in complex with Man(α1-3)Man were prepared as described for the methyl-α-D-glucose complex,[22] with the exception that 10 mM Man(α1-3)Man was used as a ligand instead of the glucoside. The complexes of the lectin with the other carbohydrates given in Table 1 were obtained by transferring the crystals of the Man(α1-3)Man complex to artificial mother liquor (200 mM calcium acetate, 100 mM sodium cacodylate (pH 6.5), 20% (w/v) PEG8000) containing increasing concentrations of the desired ligand (100 mM final concentration, reached in four steps).

All X-ray data were collected at room temperature on the EMBL beamlines of the DESY synchrotron (Hamburg, Germany). The data were processed with DENZO and SCALEPACK.[39] The statistics of the data collections are given in Table 1.

Structure determination

The structure of the Man(α1-3)Man complex was determined by molecular replacement using the co-ordinates of lentil lectin (pdb code 1len[40]) as a search model. Two clear solutions were found with AMoRe[41] that together constructed the lectin dimer. Refinement was carried out using the mlf target of CNS 1.0.[42] Cross-validation, bulk solvent correction and anisotropic temperature factor scaling were used throughout. Rounds of slow-cool simulated annealing and restrained *B*-factor refinement using all available data were alternated with manual fitting in electron density maps using TURBO.[43] At the beginning of the study, the amino acid sequence was still unknown and a sequence was derived directly from the electron density maps. At the end of the refinement, the cDNA sequence became available, and this information was used to finalise the structure. At this stage, simulated annealing was abandoned in favour of conventional positional refinement and water molecules were fit into the electron density.

The structures of all other isomorphous lectin–carbohydrate complexes were determined using the refined co-ordinates of the Man(α1-3)Man complex (stripped from its carbohydrate ligands and water molecules) as the starting model. After rigid body refinement, a slow-cool stage was used to uncouple *R* and R_{free}. From then on, restrained positional and *B*-factor refinement were alternated with manual fitting in electron density maps. The refinement statistics for all complexes are given in Table 1.

Superpositions of crystal structures were done using TURBO.[43] Figures 2–10 were produced using MOLSCRIPT[44] and Raster3D.[45]

Data Bank accession numbers

The nucleotide sequences of partial and full-length cDNAs have been deposited at GenBank with accession numbers from AJ426054 to AJ426062.

Co-ordinates and structure factors were deposited at the RCSB Protein Data Bank with as entries 1ukg, 1q8o, 1q8p, 1q8q, 1q8s and 1q8v.

Acknowledgements

This work was supported by the Vlaams Interuniversitair Instituut voor Biotechnologie (VIB), the Onderzoeksraad of the V.U.B. (grants OZR846, OZR727 and OZR728), the Fonds voor Wetenschappelijk Onderzoek Vlaanderen (FWO) and the Directorate General of International Co-operation (DGIC). The authors acknowledge the use of synchrotron beamtime at the EMBL beamlines at the DORIS storage ring (Hamburg, Germany), and the Zimbabwe Forestry Commission (Seed Research Centre, Harare) for providing the seeds. W. Versées and S. De Vos collected the X-ray data on the Man(α1-6)Man complex. R.L. & J.B. are postdoctoral fellows of the FWO. I.V.W. received financial support from the IWT.

References

1. Taylor, M. E. & Drickamer, K. (2003). *Introduction to Glycobiology*, Oxford University Press, Oxford, UK.
2. Sharma, V. & Surolia, A. (1997). Analyses of carbohydrate recognition by legume lectins: size of the combining site loops and their primary specificity. *J. Mol. Biol.* 267, 433–445.
3. Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. (1998). Legume lectin structure. *Biochim. Biophys. Acta*, 1383, 9–36.
4. Bouckaert, J., Loris, R., Poortmans, F. & Wyns, L. (1995). The crystallographic structure of metal-free concanavalin A at 2.5 Å resolution. *Proteins: Struct. Funct. Genet.* 23, 510–540.
5. Bouckaert, J., Dewallef, Y., Poortmans, F., Wyns, L. & Loris, R. (2000). Structural dissection of the conformational pathways of de- and re-metallization of concanavalin A. *J. Biol. Chem.* 275, 19778–19787.
6. Lescar, J., Loris, R., Mitchell, E., Gautier, C., Chazalet, V., Cox, V. *et al.* (2002). Isolectins I-A and I-B of *Griffonia (Bandeiria) simplicifolia*: crystal structure of metal-free GSI-B4 and molecular bases for metal binding and monosaccharide specificity. *J. Biol. Chem.* 277, 6608–6614.
7. Manoj, N., Srinivas, V. R., Surolia, A., Vijayan, M. & Suguna, K. (2000). Carbohydrate specificity and salt-bridge mediated conformational change in acidic winged bean agglutinin. *J. Mol. Biol.* 302, 1129–1137.
8. Prabu, M. M., Sankaranarayanan, R., Puri, K. D., Sharma, V., Surolia, A., Vijayan, M. & Suguna, K. (1998). Carbohydrate specificity and quaternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å resolution. *J. Mol. Biol.* 276, 787–796.
9. Audette, G. F., Vandonselaar, M. & Delbaere, L. T. (2000). The 2.2 Å resolution structure of the O(H) blood-group-specific lectin I from *Ulex europaeus*. *J. Mol. Biol.* 304, 423–433.
10. Rabijns, A., Verboven, C., Rouge, P., Barre, A., Van Damme, E. J., Peumans, W. J. & De Ranter, C. J. (2001). Structure of a legume lectin from the bark of

Robinia pseudoacacia and its complex with *N*-acetylgalactosamine. *Proteins: Struct. Funct. Genet.* **44**, 470–478.

11. Tempel, W., Tschampel, S. & Woods, R. J. (2002). The xenograft antigen bound to *Griffonia simplicifolia* lectin 1-B(4). X-ray crystal structure of the complex and molecular dynamics characterization of the binding site. *J. Biol. Chem.* **277**, 6615–6621.
12. Svensson, C., Teneberg, S., Nilsson, C. L., Kjellberg, A., Schwarz, F. P., Sharon, N., Krengel, U. (2002). High-resolution crystal structures of *Erythrina cristagalli* lectin in complex with lactose and 2′-alpha-L-fucosyllactose and correlation with thermodynamic binding data. *J. Mol. Biol.* **321**, 69–83.
13. Hamelryck, T. W., Loris, R., Bouckaert, J., Dao-Thi, M. H., Strecker, G., Imberty, A. *et al.* (1999). Carbohydrate binding, quaternary structure and a novel hydrophobic binding site in two legume lectin oligomers from *Dolichos biflorus*. *J. Mol. Biol.* **286**, 1161–1177.
14. Buts, L., Dao-Thi, M. H., Loris, R., Wyns, L., Etzler, M. & Hamelryck, T. (2001). Weak protein–protein interactions in lectins: the crystal structure of a vegetative lectin from the legume *Dolichos biflorus*. *J. Mol. Biol.* **309**, 193–201.
15. Hamelryck, T. W., Moore, J. G., Chrispeels, M. J., Loris, R. & Wyns, L. (2000). The role of weak protein–protein interactions in multivalent lectin–carbohydrate binding: crystal structure of cross-linked FRIL. *J. Mol. Biol.* **299**, 875–883.
16. Sanz-Aparicio, J., Hermoso, J., Grangeiro, T. B., Calvete, J. J. & Cavada, B. S. (1997). The crystal structure of *Canavalia brasiliensis* lectin suggests a correlation between its quaternary conformation and its distinct biological properties from concanavalin A. *FEBS Letters*, **405**, 114–118.
17. Wah, D. A., Romero, A., Gallego del Sol, F., Cavada, B. S., Ramos, M. V., Grangeiro, T. B. *et al.* (2001). Crystal structure of native and Cd/Cd-substituted *Dioclea guianensis* seed lectin. A novel manganese-binding site and structural basis of dimer-tetramer association. *J. Mol. Biol.* **310**, 885–894.
18. Imberty, A., Gautier, C., Lescar, J., Perez, S., Wyns, L. & Loris, R. (2000). An unusual carbohydrate binding site revealed by the structures of two *Maackia amurensis* lectins complexed with sialic acid-containing oligosaccharides. *J. Biol. Chem.* **275**, 17541–17548.
19. Loris, R., De Greve, H., Dao-Thi, M.-H., Messens, J., Imberty, A. & Wyns, L. (2000). Structural basis of carbohydrate recognition by lectin II from *Ulex europaeus*, a protein with a promiscuous carbohydrate-binding site. *J. Mol. Biol.* **301**, 987–1002.
20. Rozwarski, D. A., Swami, B. M., Brewer, C. F. & Sacchettini, J. C. (1998). Crystal structure of the lectin from *Dioclea grandiflora* complexed with core trimannoside of asparagine-linked carbohydrates. *J. Biol. Chem.* **273**, 32818–32825.
21. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **10**, 1–6.
22. Loris, R., Imberty, A., Beeckmans, S., Van Driessche, E., Read, J. S., Bouckaert, J. *et al.* (2003). Crystal structure of *Pterocarpus angolensis* lectin in complex with glucose, sucrose and turanose. *J. Biol. Chem.* **278**, 16297–16303.
23. Chandra, N. R., Prabu, M. M., Suguna, K. & Vijayan, M. (2001). Structural similarity and functional diversity in proteins containing the legume lectin fold. *Protein Eng.* **14**, 857–866.
24. Manoj, N. & Suguna, K. (2001). Signature of quaternary structure in the sequences of legume lectins. *Protein Eng.* **14**, 735–745.
25. Srinivas, V. R., Reddy, G. B., Ahmad, N., Swaminathan, C. P., Mitra, N. & Surolia, A. (2001). Legume lectin family, the "natural mutants of the quaternary state", provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys. Acta*, **1527**, 102–111.
26. Derewenda, Z., Yariv, J., Helliwell, J. R., Kalb, A. J., Dodson, E. J., Papiz, M. Z. *et al.* (1989). The structure of the saccharide-binding site of concanavalin A. *EMBO J.* **8**, 2189–2193.
27. Bourne, Y., Roussel, A., Frey, M., Rouge, P., Fontecilla-Camps, J. C. & Cambillau, C. (1990). Three-dimensional structures of complexes of *Lathyrus ochrus* isolectin I with glucose and mannose: fine specificity of the monosaccharide-binding site. *Proteins: Struct. Funct. Genet.* **8**, 365–376.
28. Rini, J. M., Hardman, K. D., Einspahr, H., Suddath, F. L. & Carver, J. P. (1993). X-ray crystal structure of a pea lectin-trimannoside complex at 2.6 Å resolution. *J. Biol. Chem.* **268**, 10126–10132.
29. Delbaere, L. T., Vandonselaar, M., Prasad, L., Quail, J. W., Wilson, K. S. & Dauter, Z. (1993). Structures of the lectin IV of *Griffonia simplicifolia* and its complex with the Lewis b human blood group determinant at 2.0 Å resolution. *J. Mol. Biol.* **230**, 950–965.
30. Moothoo, D. N., Canan, B., Field, R. A. & Naismith, J. H. (1999). Man(α1-2)Manα-OMe-concanavalin A complex reveals a balance of forces involved in carbohydrate recognition. *Glycobiology*, **9**, 539–545.
31. Bouckaert, J., Loris, R., Hamelryck, T. & Wyns, L. (1999). The crystal structures of Man(α1-3)Man and Man(α1-6)Man complexed to concanavalin A. *J. Biol. Chem.* **274**, 29188–29195.
32. Brewer, C. F. & Brown, R. D. (1979). Mechanism of binding of mono- and oligosaccharides to concanavalin A: a solvent proton magnetic relaxation dispersion study. *Biochemistry*, **18**, 2555–2562.
33. Brewer, C. F. & Bhattacharyya, L. (1986). Specificity of concanavalin A binding to asparagine-linked glycopeptides. A nuclear magnetic relaxation dispersion study. *J. Biol. Chem.* **261**, 7306–7310.
34. Mandal, D. K., Kishore, N. & Brewer, C. F. (1994). Thermodynamics of lectin-carbohydrate interactions. Titration microcalorimetry measurements of the binding of N-linked carbohydrates and ovalbumin to concanavalin A. *Biochemistry*, **33**, 1149–1156.
35. Naismith, J. H. & Field, R. A. (1996). Structural basis of trimannoside recognition by concanavalin A. *J. Biol. Chem.* **271**, 972–976.
36. Loris, R., Maes, D., Poortmans, F., Wyns, L. & Bouckaert, J. (1996). A structure of the complex between concanavalin A and methyl-3,6-di-*O*-(alpha-D-mannopyranosyl)-alpha-D-mannopyranoside reveals two binding modes. *J. Biol. Chem.* **271**, 30614–30618.
37. Frohman, M. A. (1993). Rapid amplification of complementary DNA ends for generation of full-length complementary DNAs: thermal RACE. *Methods Enzymol.* **218**, 340–356.
38. Sanger, F., Nicklen, S. & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl Acad. Sci. USA*, **74**, 5463–5467.
39. Otwinowski, Z. & Minor, W. (1997). Processing of

X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
40. Loris, R., Steyaert, J., Maes, D., Lisgarten, J., Pickersgill, R. & Wyns, L. (1993). Crystal structure determination and refinement at 2.3-Å resolution of the lentil lectin. *Biochemistry*, **32**, 8772–8781.
41. Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
42. Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.
43. Roussel, A. & Cambillau, C. (1989). TURBO-FRODO. *Silicon Graphic Geometry Partner Directory*, pp. 71–78, Silicon Graphics, Mountain View, CA.
44. Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
45. Merritt, E. A. & Bacon, D. J. (1997). Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524.

Edited by R. Huber

(Received 28 August 2003; received in revised form 18 November 2003; accepted 21 November 2003)

Protein Engineering vol.14 no.11 pp.857–866, 2001

Structural similarity and functional diversity in proteins containing the legume lectin fold

Nagasuma R.Chandra[1], M.M.Prabu[2], K.Suguna[2] and M.Vijayan[2,3]

[1]Bioinformatics Centre and [2]Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

[3]To whom correspondence should be addressed.
E-mail: mv@mbu.iisc.ernet.in

Knowledge of structural relationships in proteins is increasingly proving very useful for *in silico* characterizations and is also being exploited as a prelude to almost every investigation in functional and structural genomics. A thorough understanding of the crucial features of a fold becomes necessary to realize the full potential of such relationships. To illustrate this, structures containing the legume lectin-like fold were chosen for a detailed analysis since they exhibit a total lack of sequence similarity among themselves and also belong to diverse functional families. A comparative analysis of 15 different families containing this fold was therefore carried out, which led to the determination of the minimal structural principles or the determining region of the fold. A critical evaluation of the structural features, such as the curvature of the front sheet, the presence of the hydrophobic cores and the binding site loops, suggests that none of them are crucial for either the formation or the stability of the fold, but are required to generate diversity and specificity to particular carbohydrates. In contrast, the presence of the three sheets in a particular geometry and also their topological connectivities seem to be important. The fold has been shown to tolerate different types of protein–protein associations, most of them exhibiting different types of quaternary associations and some even existing as complexes with other folds. The function of every family in this study is discussed with respect to its fold, leading to the suggestion that this fold can be linked to carbohydrate recognition in general.

Keywords: carbohydrate binding/β-sandwich/structural determinants/structural relationships

Introduction

Examination of the hitherto identified protein folds reveals that the available protein structures cluster into limited regions of the entire conformational space (Holm and Sander, 1995). This means that several protein families share a common structural fold, some of which are obvious from the sequence similarities that they exhibit. It is also well known that protein evolution gives rise to families of structurally related proteins, within which sequence similarities can be extremely low. Such unanticipated relationships in known structures have been identified effectively by structure-based classifications (Holm and Sander, 1996). Several excellent databases featuring structural classifications of protein structures have been developed in recent years [SCOP (Murzin *et al.*, 1995), FSSP (Holm and Sander, 1996), CATH (Orengo *et al.*, 1997)]. These databases serve as useful guidelines to study the overall folds and structures of various proteins encoded by a genome. To realize the full potential of these relationships, it is essential to characterize the structural determinants or the minimal structural principles of the individual folds. The completion of several genome sequences including that of the human genome provides an additional impetus to such characterizations. Structures containing the legume lectin-like fold are so diverse in the families to which they belong that they provide classical examples for investigations of this type.

Lectins are carbohydrate-binding proteins that specifically recognize diverse sugar structures and mediate a variety of biological processes such as cell–cell and host–pathogen interactions, serum glycoprotein turnover and innate immune responses (Vijayan and Chandra, 1999). Lectins are found in most organisms, ranging from viruses and bacteria to plants and animals (Lis and Sharon, 1998). They represent a heterogeneous group of oligomeric proteins that vary widely in size, structure, molecular organization and the constitution of their combining sites. Nonetheless, many of them belong to distinct protein families, classified based on biochemical, functional or structural properties. Although a number of lectins have been well studied, ambiguities still exist in their precise biological roles. A well-studied class of lectins from leguminous plants contain a characteristic fold (Loris *et al.*, 1998; Bouckaert *et al.*, 1999), often referred to as the legume lectin fold or simply the lectin fold. This fold is one of the widely occurring protein folds represented by 14 distinct protein families, in addition to legume lectins. The most striking feature of the fold is the total lack of sequence similarity among different members exhibiting the fold. It is remarkable indeed that different members exhibiting this fold can show as low as 2% sequence identity. The recent addition of the structures of several legume lectins and also many other proteins possessing the fold (Berman *et al.*, 2000; Bettler *et al.*, 2001) enables us to analyse and compare these various structures in order to characterize the structural features of this fold. Here, we seek to compare the structures of the 15 different families and derive common structural features and determinants of the fold. In an attempt to relate fold to function, we also analyse features required for carbohydrate recognition, which happens to be the best recognized function of members with this fold.

Methods

Initial identification of proteins containing the legume lectin-like fold was made using the SCOP (Murzin *et al.*, 1995) and FSSP databases (Holm and Sander, 1996), which was followed by an analysis of related proteins in the CATH (Orengo *et al.*, 1997) and the 3D lectin databases (Bettler *et al.*, 2001). Further, to identify structural homologues, a thorough investigation of all available protein structures in the protein data bank was carried out using two separate structure comparison algorithms,

DALI (Holm and Sander, 1995) and VAST (Gibrat *et al.*, 1996). All the identified proteins were analysed for particular features and re-classified based broadly on their known functions. Only those proteins which had a Z-score >2.5, for at least 70% of the contents of the relevant domain, from the DALI comparisons are included here. This study involved an analysis of more than 300 structures, which were obtained from our local repository of coordinates, regularly downloaded from the Protein Data Bank (Berman *et al.*, 2000). The MSI software package (InsightII, version 98) was used to visualize, analyse and manipulate various structures. The solvent accessibility calculations were carried out using the Connolly algorithm (Connolly, 1993). To deduce topology, sequential connectivity of the individual strands in the three β-sheets were considered. The curvature of the sheet was computed by measuring virtual angles subtended by the end Cα atoms at their mid-points. The average distance between two sheets was computed by measuring the distance between the centroids of each sheet using only the Cα atoms of each sheet. Hydrophobic cores were considered present, when a minimum of four residues, that had an overall accessibility of <10%, were in contact within a radius of 5 Å of each other, forming a cluster.

Results and discussion

General description of the fold

The structure of concanavalin A (ConA), the first legume lectin to be X-ray analysed, in 1972, exhibited a β-sandwich type of structure (Hardman and Ainsworth, 1972). Subsequently, about 100 more structures involving more than 15 different legume lectins in their complexed and uncomplexed forms have been studied. All of them share the same tertiary structure with that of ConA in their individual subunits. These include about 10 structures of peanut and winged bean basic and acidic lectins from our laboratory (Banerjee *et al.*, 1994, 1996; Prabu *et al.*, 1998; Manoj *et al.*, 2000). A structure-based classification of proteins places them in a super-family of ConA-like lectins and glucanases, one of the many in the all-β structural class. The subunits of legume lectins are most often made up of single polypeptide chains of ~250 amino acids exhibiting the legume lectin fold. The fold primarily consists of three β-sheets, a 'flat' six-membered 'back' β-sheet, a small 'top' β-sheet and a curved, seven-stranded 'front' β-sheet and a number of loops interconnecting the sheets as well as the strands in them (Banerjee *et al.*, 1996). In peanut lectin, for example, 110 out of 228 residues have β-structures; the remaining are in loops and β-turns, connecting the strands. Legume lectins within themselves exhibit remarkable sequence homologies and structural similarities, despite differences in sugar specificities and quaternary structures. Superpositions of Cα atoms of the β-sheets of individual subunits of legume lectins using various combinations results in root mean square deviation (r.m.s.d.) values ranging from about 0.6 to 2.0 Å.

Analysis of structures in the Protein Data Bank reveals that there are many other families of proteins which exhibit the same legume lectin subunit fold. Structural homologues were identified using DALI (Holm and Sander, 1995) and VAST algorithms (Gibrat *et al.*, 1996), compared with the SCOP and FSSP databases and reclassified based on their known functions, as shown in Table I. The first legume lectin to be X-ray analysed, ConA, is somewhat atypical as the post-translational modification involving a circular permutation results in a mature protein with amino and carboxyl termini at locations different from those in all other legume lectins of known three-dimensional structure. Therefore, instead of ConA, a subunit of tetrameric peanut lectin (2PEL:A), another thoroughly characterized lectin, will be used as a representative of legume lectins in the present study. The highest resolution structure of either the carbohydrate complex where available or of the apo protein from each of the other families was chosen as a representative structure for further analysis. The representative structures chosen are listed in Table I and illustrated in Figure 1. Tumour necrosis factor-α and the viral capsid proteins were also among the structural homologues of peanut lectin. They were not considered in this study, however, since their similarities lie below the cutoff criteria chosen here.

The terms β-sandwich fold and the jelly roll fold have also been used in the literature to describe these proteins. Whereas the former term is technically correct, since the legume lectin fold is a type of β-sandwich, use of the latter term is debatable as the legume lectin fold does not strictly conform to the definition of a jelly roll (Chelvanayagam *et al.*, 1992). There are minor variations in topology among the different families of proteins exhibiting this fold. Therefore, strictly, they may be described as belonging to a set of closely related folds. However, as explained later, the differences in topology are very small and the term legume lectin fold will be used to encompass all of them.

Characteristic features of the legume lectin fold

The two main β-sheets, their relative orientation and the third β-sheet. Superposition of the β-strands in all structures with those in the first subunit of peanut lectin reveals that all of them have two β-sheets with a roughly similar mutual orientation. R.m.s.ds, number of residues aligned and the structural similarity score are shown in Table I. The 'back' sheet ranged between four and six strands with seven residues in each strand, in all the structures, while the front sheet showed 5–7 strands with 5–7 residues in each strand. This variation was restricted only to the first two strands in the back sheet and the first three strands in the 'front' sheet. Strand regions corresponding to residues 64–70, 162–168, 173–179 and 186–192 from the back sheet and residues 84–90, 117–124, 136–143 and 149–153 of the peanut lectin structure can therefore be said to be invariant. Bovine spermadhesin and the insecticidal toxin, however, exhibit minor variations on this theme, since only those which correspond to the middle strands in both sheets in peanut lectin are present. It must be remembered, however, that these two families deviate the most from the typical legume lectin fold.

The two sheets are approximately parallel to each other and also situated at a distance averaging ~13 Å between the back and the front sheets as measured by computing the distances between the centroids of the Cα atoms of each sheet. The presence of the 'back' and the 'front' sheets and similarities in their relative orientation, therefore, clearly appears to be a characteristic of the legume lectin fold. Several hydrophobic residues present on both sheets have their side chains positioned between the sheets so as to form a hydrophobic cluster, which provides an important source of stability for maintaining the fold. A prominent feature of this hydrophobic cluster is the aromatic side chains of one sheet stacking against those of the other sheet. The observed distance of ~13 Å, can be justified in terms of the optimal distance required for such interactions.

The third, small 'top' anti-parallel sheet made of five strands with only 2–4 residues in each strand exists in all legume

Table 1. Similarities in structures containing the legume lectin-like fold*

Protein name	PDB code	Z-score	Sequence length aligned	Total sequence length	R.m.s.d.	% Identity	Family
Peanut lectin	2PEL	–	–	224	–	–	Legume lectins
α-Amylase inhibitor	1DHK(B)	16.2	118	195	1.3	44	Lectin-like inhibitors of amylase
Arcelin-5	1IOA	15.8	121	228	1.3	36	Phytohaemagglutinins
Glucanase	2AYH	10.6	112	214	2.2	13	Glucanases
Tetanus toxin C fragment	1DLL	9.5	161	452	3.6	12	Neurotoxins
Charcot Leyden crystal protein	1QKQ	8.9	95	141	2.5	4	Lysophospholipases
Neuraminidase	1KIT	8.7	116	194	3.1	9	Neuraminidases/sialidases
Neurexins	1QU0	7.5	122	181	3.0	16	LNS domains
Galectin-1	1SLC	7.0	92	134	2.7	4	Galectins
C-reactive protein	1GNH	6.7	100	126	3.0	11	Acute phase proteins
Serum amyloid component	1SAC	6.4	100	204	3.0	10	Amyloid binding proteins
Cellobiohydrolase	1CEL	5.3	106	434	2.9	8	Cellulases
Xylanase-II	1RED	4.9	121	190	3.3	11	Xylanases
Insecticidal toxin	1CIY	4.7	109	577	3.6	8	Insecticidal crystal proteins
Spermadhesin	1SPP	2.8	81	111	3.8	5	Seminal fluid proteins

*Structural similarity is expressed in terms of the normalized Z-scores that refer to the strength of structural similarity in standard deviations above the mean (Holm and Sander, 1996).

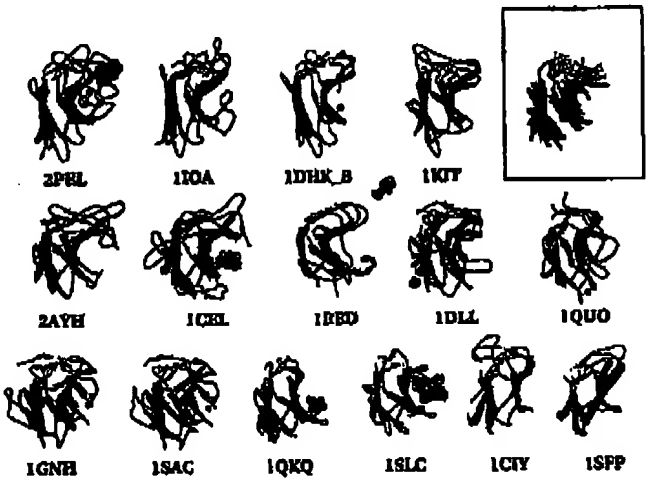


Fig. 1. Ribbon diagrams of the 15 representative structures containing the legume lectin-like fold chosen in this study. All the structures are shown in the same orientation. The PDB codes are given below each structure. The inset shows the superposition of the three sheets in all the structures. Figures 1, 2A and 4 were prepared using MOLSCRIPT (Kraulis, 1991). The back sheets are shown in black, the front sheets in dark grey and the top sheets in light grey. Carbohydrate ligands and metals where present are shown as CPK models.

lectins. The strands correspond to residues 25–27, 31–34, 217–220, 71–74 and 160–161 in peanut lectin (Banerjee *et al.*, 1996). This sheet is not explicitly acknowledged as a separate β-sheet in most of the other structures. However, upon examination of their backbone dihedral angles and hydrogen bonding patterns, all the 15 representative structures described here were found to contain segments corresponding to this sheet in an orientation analogous to that observed in peanut lectin. As in the case of the 'back' and the 'front' sheets, here too only the last three strands corresponding to residues 217–220, 71–74 and 160–161 were invariant. A schematic diagram of the fold and a stereo view of the ribbon diagram of a subunit of peanut lectin are shown in Figure 2.

Concavity of the front sheet. The front sheet is curved in all legume lectins. In order to determine how important this curvature was for determining the legume lectin fold, the extent of curvature in all 15 structures was analysed by measuring the distances between the ends of the two middle strands and the virtual angles the ends subtend at the mid-points of the strands. The end-distances and virtual angles average ~22 Å and ~150°, respectively for galectins, Charcot Leyden crystal protein, spermadhesins and the insecticidal toxin, while the average values for legume lectins, arcelins, cellobiohydrolase, glucanases, xylanases, neurexins and tetanus neurotoxin are ~19 Å and ~120°. The pentraxins show values between these two types, with one of the strands showing more curvature than the other. These calculations confirm that whereas the front sheet is almost flat in some proteins, it is significantly curved in most others, giving rise to a concave surface, suggesting that the fold can tolerate considerable

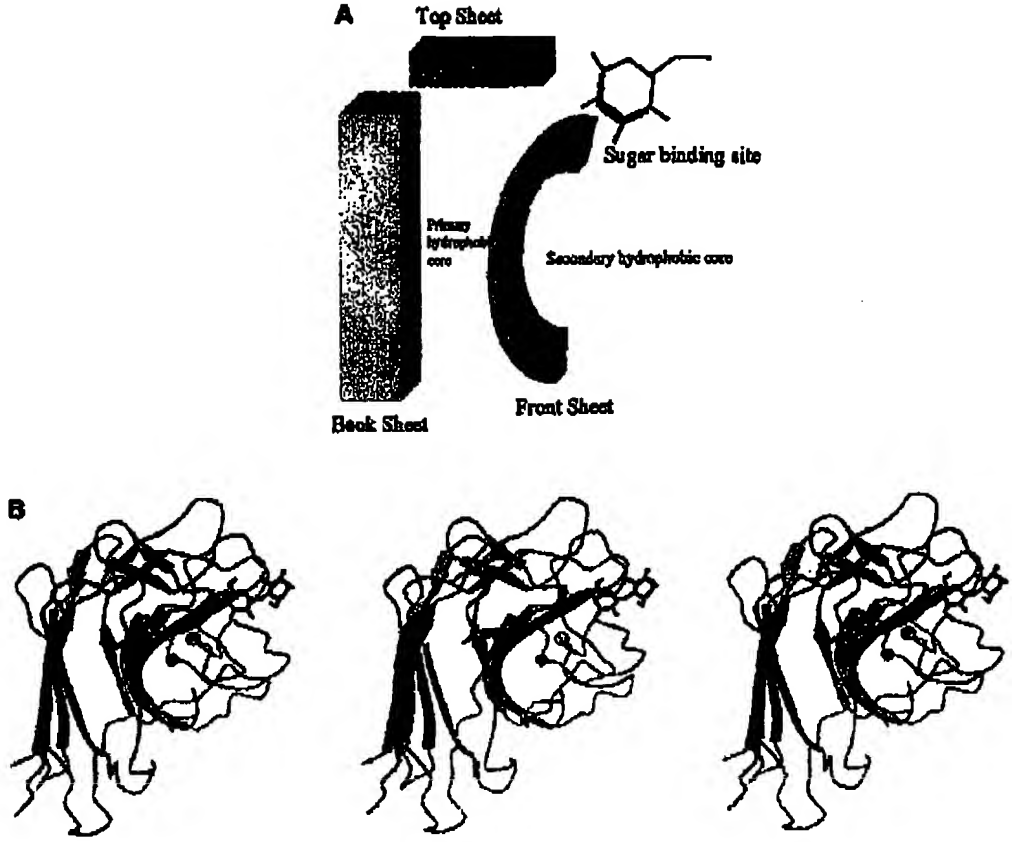


Fig. 2. (A) Schematic diagram of the legume lectin-like fold showing the position of the three sheets and the carbohydrate in legume lectins. (B) Stereo view of the ribbon diagram of a subunit of peanut lectin. The colour scheme used is similar to that in Figure 1.

variation in this parameter (Table II). It also suggests that the curvature of the front sheet is not too critical for either the formation or the stability of the overall fold.

Hydrophobic cores and surface loops. In addition to the first hydrophobic core between the two main β-sheets, present in all structures, a second hydrophobic patch has also been observed in some structures such as peanut lectin and cellobiohydrolase. Upon examination of these structures, it appears that along with the spatial disposition of long loops connecting the strands, the curvature of the front sheet is responsible for this. Indeed, only those structures that had a curved front sheet and large loops connecting the strands in it exhibited a second hydrophobic patch and those structures with either a nearly flat front sheet or with curved sheets with short loops did not exhibit this phenomenon (Table II). The solvent-accessible surface areas computed for various structures confirm the presence of the first hydrophobic patch in all structures and the presence of the second patch in some, correlating well with the curvature of the front sheet and the presence of large loops. It therefore appears that the second hydrophobic patch is not crucial for the fold. Each family showed a number of varying loops in their member structures and in many cases loops played an important role in carbohydrate binding, e.g. the four large loops in legume lectins. Galectins, on the other hand, demonstrate that the loops can be rather small, as small as just two residues, and yet be capable of carbohydrate binding, as discussed in a later section. Again, this serves to prove that the loops do not play a critical role for the formation, stability or function of the fold but may give rise to specificity or differences in affinity for various oligosaccharides. An examination of the distribution of charged residues on the structures reveals that in most cases they are clustered on the front sheet irrespective of its extent of curvature. Charged residues were found in some structures either on the back sheet or on the top sheet, but were often explainable in terms of their quaternary associations.

Topology. The legume lectin fold has three β-sheets, as has been described in the previous section, described in SCOP as exhibiting complex topology. The fold contains primarily a β-sandwich made of strands (1, 18, 6, 13, 14, 15)–(2, 5, 16, 8, 9, 10, 11) with a lid made of strands (3, 4, 17, 7, 12) where each number denotes that assigned to the strand based on their position in the sequence. The pairs formed within the fold are all antiparallel and made of strands 1–18, 2–5, 3–4, 4–17, 5–16, 6–13, 6–18, 7–12,7–17, 8–9, 8–16, 9–10, 10–11, 13–14 and 14–15. The connectivities of the strands in the three sheets

Table II. Descriptors of the fold[a]

PDB code	Concavity	Hydrophobic core		Loops in binding site	Quaternary structure	Other quaternary types in the family
		1	2			
2PEL	+	+	+	4	Tetramer	Various types of dimers and tetramers
1DHK(B)	+	+	+	0	Dimer	Bound to α-amylase
1IOA	+	+	+	3	Monomer	Dimer
2AYH	+	+	-	0	Monomer	Dimer
1DLL	+	+	+	3	Domain of a larger protein	-
1QKQ	-	+	-	2 short	Dimer	-
1KIT	+	+	-	Short	Domain of a larger protein	-
1QU0	+	+	+	3	Monomer	Various multimers
1SLC	-	+	-	2	Dimer	Monomer
1GNH	+	+	-	2	Pentamer	-
1SAC	+	+	-	2	Pentamer	-
1CEL	+	+	+	Extra domain	Monomer	-
1RED	+	+	+	3	Monomer	Dimer
1CIY	-	+	-	0	Domain of a larger protein	-
1SPP	-	+	-	0	Heterodimer	Homodimer

[a]Concavity is described as + when the angle defining it (see text) is < 135°.

were identical, except for the small differences indicated below, among the superposable parts of all the representative structures, suggesting that topology is an important characteristic feature of the fold. Figure 3 depicts the topological connections in peanut lectin in a schematic diagram. The figure also illustrates how this topology is substantially conserved in all the representative structures in spite of variations in the fold itself. The variations are primarily truncations of the sheets, insertions of additional segments or changes in the positions of the N- and C-termini in the sequence. A classical example is provided by ConA, where the topology of the sheets remains identical with that of peanut lectin except that the N- and C-termini are at different positions. The same phenomenon is observed in the case of both the pentraxins, where the termini are merely frame-shifted by three strands in the sequence. Pentraxins also have two α-helices, one between the first two strands of the back β-sheet and the other between the front and the top sheets. An insertion is observed in the wings of neuraminidase also, where an α-helix is observed between the back and the front sheets, without any changes in either the position of the termini or the topology. The structure of cellobiohydrolase presents a good example of an entire domain being inserted between the back and the front sheets and two smaller domains, predominantly consisting of α-helices, between the front and the top sheets. Xylanases reveal an insertion of additional strands in its front sheet leading to two additional β-hairpins. Galectins and the Charcot Leyden protein demonstrate yet another variation, that can be described as a truncation of the first four strands in the sequence resulting in one strand less in both the main sheets, as compared with the situation in peanut lectin. Spermadhesins and the insecticidal toxin show larger truncations within the back and the front sheets, but again the topologies of the existing strands remain similar. These observations suggest that the fold can tolerate insertions and deletions such as those described, without any disturbance to its overall nature.

Derivation of minimal structural determinants. Compilation of the residues in the superimposable segments in all structures reveals that they are all close together in their sequential positions, indeed clustered into one long segment of the polypeptide, corresponding to residues 64–192 of peanut lectin. This segment corresponds to a β-hairpin in the front sheet, a double hairpin of the back sheet and two short strands on the top sheet which are basically extensions of two invariant strands in the back sheet. Interestingly, all these regions are contained within one contiguous segment of the polypeptide chain in all the structures studied here. This segment, highlighted in Figure 3, can be described as the minimal structural principle or as the determining region for the legume lectin fold. The minor topological variations observed in some proteins, especially with reference to the positions of their chain termini (e.g. in ConA or the pentraxins), occur between the first and the second strands of the back sheet and do not alter the invariant region in the fold, consistent with the hypothesis that the invariant region is indeed the determining region. The only variation to this general consensus is observed in the structures of the insecticidal toxin and spermadhesin where a segment corresponding to two strands in the back and front sheets within this determining region are deleted, but there too, the remaining strands of the invariant region are within one sequential segment.

Quaternary structure

Legume lectins themselves exhibit different types of quaternary structures not only in terms of the number of subunits involved in the oligomeric molecule but also in the nature or type of oligomerization, despite having nearly identical subunit structures. The dimerization of ConA and several other legume lectins involves the association of the two back β-sheets into a contiguous 12-stranded β-sheet in a side-by-side arrangement, with the dyad of the dimer perpendicular to the β-sheet thus formed. That in peanut lectin, lectin IV of *Griffonia simplicifolia*, *Erythrina corallodendron* lectin and winged bean lectins involves the back-to-back arrangements of the two 'back' β-sheets. All tetrameric legume lectins are dimers of

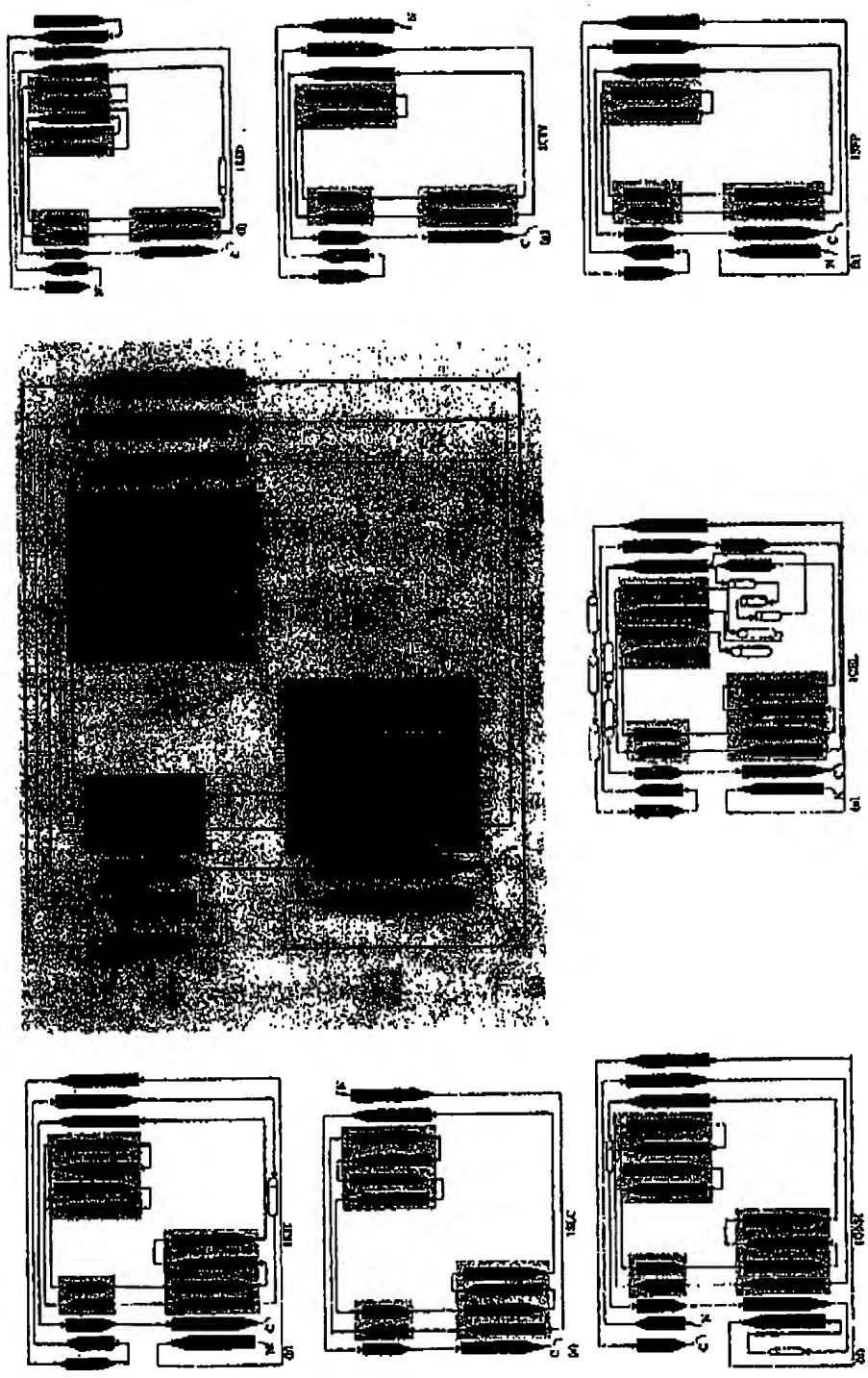


Fig. 3. Topology diagram of (a) peanut lectin shown in the centre panel and its subtle variations in other structures: (b) neuraminidase, (c) galectin, (d) pentraxin, (e) cellobiohydrolase, (f) xylanase, (g) insecticidal toxin and (h) spermadhesin. The topology diagrams of arcelins, α-amylase inhibitor, tetanus neurotoxin, glucanases and the LNS domains are similar to that of peanut lectin, that of Charcot Leyden crystal protein is similar to that of galectins, while both the pentraxins have the same topological type. The invariant regions in each of them are highlighted in grey. Strands are numbered according to their three-dimensional positions in the respective β-sheets in peanut lectin. Strands in the other structures are numbered with reference to those in peanut lectin. Asterisks denote those strands that do not match with any strand in peanut lectin.

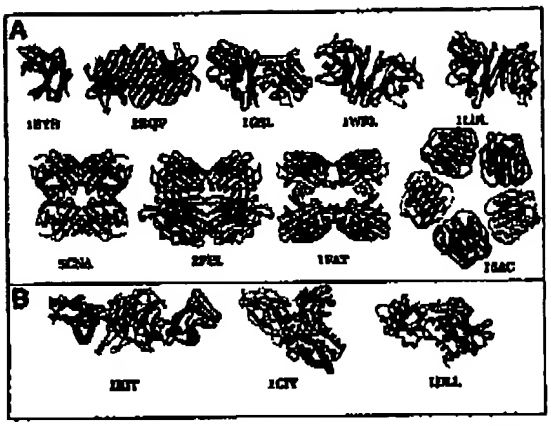


Fig. 4. (A) Various types of quaternary associations exhibited by members of the legume lectin-like fold. (B) Examples where legume lectin-like folds exist as domains of multidomain proteins, as in neuraminidase (1KIT), insecticidal toxin CryIIIA (1CIY) and the tetanus neurotoxin (1DLL).

dimers, but here again considerable variability exists (Prabu *et al.*, 1999). In this context, the most interesting case is presented by peanut lectin in which the tetrameric molecule has an 'open' structure without the expected 222 or 4-fold symmetry (Banerjee *et al.*, 1994). The ConA type of dimerization is exhibited by galectins, sialidase, a member of the neuraminidase family, arcelin-1 and the lectin-like inhibitor of amylase, while a back-to-back arrangement is found in the porcine seminal plasma spermadhesin heterodimer. The pentameric pentraxins exhibit yet another type of oligomerization. The association of the A–B subunits appears rather weak although involving the back β-sheets; the association of the A–B subunits involves only the top β-sheet and some loops around it, thus forming a new mode of dimerization. It is worth mentioning that the invariant region is not associated with quaternary association in any of the structures studied here. Figure 4 illustrates that structures containing the legume lectin-like fold exist as monomers, different types of dimers, tetramers and also as domains of multi-domain proteins. Differences in quaternary associations have been shown to give rise to differences in oligosaccharide specificities in bulb lectins (Chandra *et al.*, 1999), although any such clear correlations have not been discovered so far for the proteins used in this study.

Biological role and carbohydrate binding

The main criterion for a protein to be classified as a lectin is its ability to bind a carbohydrate, most often an oligosaccharide (Lis and Sharon, 1998). Lectins are known to mediate a variety of cellular interactions through their ability to bind carbohydrates specifically. It is not surprising, therefore, that different lectins are specific to different carbohydrates. A study of the crystal structures of more than 50 legume lectin–carbohydrate complexes has shown that the lectins bind the carbohydrates at the top of the concave side of the front sheet and involve interactions from the four loops (91–106, 125–135, 75–83 and 211–216). The first three loops are largely common to all legume lectins whereas the fourth one varies in size and conformation and is thought to determine specificity of the lectin. The predominant function of legume lectins, all exhibiting the same fold, is therefore carbohydrate binding (Sharma and Surolia, 1997). The obvious questions that this observation raises are whether the legume lectin-like fold is always involved in carbohydrate binding and whether the fold gives rise to that particular function. A brief description of the families of proteins in this study provides some insight into these questions.

Galectins, represented by 1SLC, are a family of soluble animal lectins that are cation dependent and bind to Gal-β(1,4)GlcNAc terminating oligosaccharides (Bourne *et al.*, 1994). They have been implicated in modulation of cell–cell interactions through carbohydrate-mediated recognition. The crystal structures of five galectins reported so far all show that the carbohydrate binds on the front sheet. The front sheet, however, is not curved in galectins and the loops corresponding to the four carbohydrate binding loops on peanut lectin are either very small or are not present. Carbohydrate binding is achieved through interactions of the charged and polar residues on the front sheet, especially an aspartic acid, an asparagine and an arginine. It appears that the longer side chains seen on the front sheet compensate for the loss of the loop regions and succeed in binding the carbohydrate in a somewhat similar position. The human Charcot Leyden crystal protein, also referred to as galectin-10 owing to its structural similarity to galectins, is the major autocrystallizing constituent of human eosinophils and basophils during allergic inflammation and is known to possess lysophospholipase activity (Swaminathan *et al.*, 1999). The CLC protein structure possesses a carbohydrate recognition site comprising most of the binding residues that are conserved among galectins. The protein exhibits specific although weak, binding to mannose, *N*-acetylglucosamine and lactose. The binding site of mannose in the crystal structure of a complex is seen to be similar to the sugar binding site in galectins.

Arcelin-1, a member of the phytohaemagglutinin family, is a glycoprotein from kidney beans (*Phaseolus vulgaris*) which displays insecticidal properties and protects the seeds from predation by larvae of various bruchids (Mourey *et al.*, 1998). This lectin-like protein, although devoid of monosaccharide binding properties, exhibits specificity for various glycoproteins such as fetuin and asialofetuin. The related protein arcelin-5 has a different quaternary structure, binds to monosaccharides specifically, on the concave side of the front sheet involving interactions similar to those observed in legume lectins (Hamelryck *et al.*, 1996). The differences in function between the two arcelins, due to the differences in their carbohydrate-binding properties, have been explained in terms of sequence and structural changes in the two proteins. The seeds of *Phaseolus vulgaris* contain another protein that inhibits α-amylase in the digestive tract of mammals and coleoptera and the growth of burchid larvae (Bompard-Gilles *et al.*, 1996). The structure of this compound, determined a few years ago, reveals a lectin-like domain with a tertiary structure very similar to that of ConA. The carbohydrate-binding loops in this protein are truncated, facilitating its binding to amylase.

Pentraxins are pentameric plasma glycoproteins characterized by calcium-dependent ligand binding. Their overall structures have been described earlier to be similar to that of legume lectins (Srinivasan *et al.*, 1996). The loops within and between the β-sheets are much shorter in this family of proteins. The human serum amyloid P component binds to 4,6-cyclic pyruvate acetal of β-D-galactose and all forms of amyloid fibrils through calcium ions (Emsley *et al.*, 1994). Although the positions of calcium ions are different from the

metal ion positions in legume lectins, the mode of recognition bears some resemblance in the two families, especially the role played by Gln148 in the serum amyloid component, as compared with that of Asn127 in peanut lectin. The human C-reactive protein, although belonging to the same structural class as that of the serum amyloid component, is functionally a completely different protein (Shrive *et al.*, 1996). It is an acute phase reactant protein that is expressed rapidly as a response to infection or injury and is known to be involved in enhancement of phagocytosis and activation of the complement through its ability to bind to the bacterial polysaccharides. The ligand is expected to bind at the concave side of the front sheet in both members of the pentraxin family.

Many microorganisms produce multiple forms of different 1,4-β-glycosidases in order to hydrolyse plant polysaccharides such as cellulose and xylan. Because of the complex nature of these polysaccharides, different glycosidases are required to complete their hydrolysis. These glycosidases are classified into *endo-* or *exo*-enzymes depending on their ability to catalyse the backbone breakage of their polymeric substrate (Wood, 1989). Microbial cellobiohydrolase is a representative of the exoglucanase/cellulase family, evolved to carry out the hydrolysis of cellulose, the major polysaccharide in plants (Divne *et al.*, 1994). Its crystal structure shows that this protein specifically binds to the 1,4-α-D-glucan moiety of cellulose. The glucanases (Keitel *et al.*, 1993), also known as lichenases, represent a distinct family of glucanohydrolases, that primarily hydrolyse the disordered amorphous regions of cellulose, cutting at internal glycosidic bonds. The two proteins together act in synergy to achieve complete hydrolysis of cellulose. Although there are several differences in the two structures, especially in terms of insertions and the number of loops, both appear to bind carbohydrate moieties on the concave side of the front sheet, involving interactions of residues structurally related to those in legume lectins. In the case of cellobiohydrolase, the extra loop regions and insertions serve to form a tunnel to encompass the product cellobiose (Divne *et al.*, 1994). Xylanases, another class of endoglycosidases, are structurally very similar to glucanases and have their binding sites situated in a cleft on the concave side of the front sheet. In this class of proteins, clear conformational changes have also been observed upon ligand binding (Havaukainen *et al.*, 1996). These changes involve the extra insertions present in the front sheet and also the loop regions located close by.

Two lectin-like domains have been observed on either side of the central sialidase domain of *Vibrio cholerae* neuraminidase (Crennel *et al.*, 1994). Neuraminidase cleaves the glycosidic linkage between a terminal sialic acid and the penultimate sugar in various glycoconjugates. The environment of the small intestine requires the bacteria to secrete several adhesins. It is expected that the lectin-like domains mediate the protein's attachment to the adhesins. The ability to recognize carbohydrates by these domains, however, remains to be proven.

LNS domains are present in diverse proteins such as laminins, neurexins, agrins and slit (Rudenko *et al.*, 1999). The structures of the G domain of laminin A (Hohenester *et al.*, 1999), steroid binding protein (Grishkovskaya *et al.*, 2000) and neurexins (Rudenko *et al.*, 1999) have recently been determined and found to be extremely similar to each other. Neurexins are brain-specific cell surface proteins that are believed to be involved in neuron–neuron recognition and neuron–neuron adhesion. The crystal structure of the LNS domains in them responsible for this function, determined recently, reveal β-sandwich motifs with striking similarity to legume lectins. The LNS domains in agrin and laminin A are known to bind to heparin and other glycosaminoglycan components. The ligand-binding sites in these domains, however, vary from that in legume lectins and also in the different members containing this domain. This is not surprising given the broad range of ligands that these domains can bind. The structural homology between neurexin-1β and lectins raises the possibility that LNS domains may have a general function as carbohydrate-binding modules and that in neurexins, protein–carbohydrate interactions might contribute to their cell adhesive properties at neuronal junctions (Rudenko *et al.*, 1999). While it remains to be investigated whether LNS domains in neurexins do indeed bind sugars, their interactions with protein ligands such as α-latrotoxin and neuroligins are well characterized.

Tetanus neurotoxin (TeNT) is the sole causal agent of the pathological condition known as tetanus (Umland *et al.*, 1997). TeNT is a member of the clostridial neurotoxin family, the most potent toxins known. The botulinum toxin family is closely related to this. The extraordinary toxicity arises from two factors: the first is the critical importance of VAMP/synaptobrevin, the toxins' substrate to neuroexocytosis, and the second is the exquisite transport mechanism exploited by the toxin for delivery to its cytosolic target within the central nervous system. The receptor binding subunit of TeNT plays a dominant role in this delivery process, perhaps through its ability to bind carbohydrates. The structure of the neurotoxin complexed to galactose has been determined recently (Emsley *et al.*, 2000). The characteristic structural features of this protein, such as the topology, concavity of the front sheet and the presence of the binding site loops, is remarkably similar to that of legume lectins. The crystal structures of several complexes determined by Emsley *et al.* indicate that tetanus toxin has multiple carbohydrate-binding sites, the site on the domain adopting the β-trefoil fold being the best established. In addition, the β-sandwich domain also has several residues (particularly Tyr909, Glu932, Asp1067 and Asn1069) situated on the concave surface of the front sheet that appear appropriate to bind a carbohydrate molecule, analogous to that in some of the other structures discussed here. Emsley *et al.* also discuss the possibility of subsite multivalency in these proteins similar to that in lectins. Whether both domains are involved in carbohydrate recognition and, if they are, whether they are involved together or separately, however, still remain to be explored.

The crystal structure of the activated 65 kDa lepidopteran-specific CryIA and CryIIIA toxins from *Bacillus thuringiensis*, belonging to a large protein of cry proteins, reveals a domain (domain III) containing a β-sandwich structure made of two twisted antiparallel β-sheets forming a face-to-face sandwich (Grochulski *et al.*, 1995). These toxins, also known as insecticidal crystal proteins, are synthesized intracellularly as inactive prototoxins and, when activated in the gut juice, bind to high-affinity sites of the midgut epithelial cells. Analysis of the structural fold of this domain indicates it to be a variant of the jelly roll fold. The minimal determining region of the legume lectin-like fold appears intact in these proteins also, although it is significantly different from legume lectins because of differences in topology, a smaller number of strands in each sheet and loss of concavity on the front sheet and loss of equivalent binding site loops. The involvement of this domain in receptor recognition has been suggested (Grochulski *et al.*, 1995), although its exact role remains to be identified.

Spermadhesins are a family of conserved proteins known to be important in gamete recognition during fertilization (Romero *et al.*, 1997). Several members of this family have been studied of which the crystal structures of seminal plasma proteins 1 and 2 (PSP-I and PSP-II) (Varela *et al.*, 1997) and the acidic seminal plasma protein (aSFP) (Romao *et al.*, 1997) are available. Although all members of the spermadhesin family share 60–98% amino acid sequence, they are not functionally equivalent. For example, the porcine spermadhesin AWN and its equine homologue HSP-7 display carbohydrate-binding activity through which they bind tightly to the sperm head membrane, whereas the bovine aSFP does not show carbohydrate-binding activity but is thought to stimulate progesterone secretion by granulosa cells. All three structures, however, reveal a CUB domain architecture which is a variant of the jelly roll motif. The three proteins have been shown to superimpose well among themselves with r.m.s.ds below 1 Å (Romero *et al.*, 1997). The putative carbohydrate-binding site in PSP-II is suggested to be located at a shallow groove on the protein surface, similar in position to that in legume lectins and galectins (Varela *et al.*, 1997). This is despite the loss of concavity and shortening of loops in the spermadhesins.

Conclusions

The comparison of 15 different families described here suggests that their overall structures have some characteristic features common to all members. The topology is remarkably conserved in all the members, suggesting it to be a strict prerequisite for the fold. The variations that occur in spermadhesin and the insecticidal toxin do not change the nature of this region. The study also highlights the fact that the fold is compatible with different quaternary structures that are formed in different families or in different members within a family. The fold appears to tolerate a wide variation in the loops, especially those corresponding to those in the carbohydrate binding site in legume lectins. The predominant function of proteins of these 15 families appears to be carbohydrate binding through which they mediate higher order biological events, although in some cases the exact nature and specificity of the sugar are yet to be determined. Variations in the binding site and the loop lengths, observed in some cases, probably tailor the different proteins for differences in ligand specificities that are required to perform a wide range of functions such as those described here and surely many more yet to be determined. The comparative analysis presented here clearly identifies the minimal determining region in the fold and how it provides a common scaffold over which local structural variations can be rendered to achieve flexibility and adaptability required for recognizing diverse carbohydrates. Recognition of such scaffolds in every fold can be useful for automating structural classifications in the future, a need that will be increasingly on the rise as the structural genomics projects begin to make headway. More importantly, these scaffolds provide discrete templates for use in fold recognition of a new protein, another acute need that has arisen out of the sequencing of several genomes.

Acknowledgements

We thank D.V.Naimaj for discussions and help in the early stages of this work and Gosu Ramachandriah for help in preparing the figures. Use of facilities at the Super Computer Education and Research Centre, the Interactive Graphics Based Molecular Modelling Facility and the Distributed Information Centre (both supported by DBT) are acknowledged. Financial assistance from DST is acknowledged.

References

Banerjee,R., Mande,S.C., Ganesh,V., Das,K., Dhanaraj,V., Mahanta,S.K., Suguna,K., Surolia,A. and Vijayan,M. (1994) *Proc. Natl Acad. Sci. USA*, 91, 227–231.

Banerjee,R., Das,K., Ravishankar,R., Suguna,K., Surolia,A. and Vijayan,M. (1996) *J. Mol. Biol.*, 259, 281–296.

Berman,H.M., Westbrook,J., Feng,Z., Gilliland,G., Bhat,T.N., Weissig,H., Shindyalov,I.N. and Bourne,P.E. (2000) *Nucleic Acids Res.*, 28, 235–242.

Beuler,E., Loris,R. and Imberty,A. (2001) *3D Lectin Data Bank* (http://www.cermav.cnrs.fr/databank/lectine).

Bompard-Gilles,C., Rousseau,P., Rouge,P. and Payan,P. (1996) *Structure*, 4, 1441–1452.

Bouckaert,J., Hamelryck,T., Wyns, L. and Loris,R. (1999) *Curr. Opin. Struct. Biol.*, 9, 572–577.

Bourne,Y., Bolgiano,B., Liao,D.I., Strecker,G., Cantau,P., Herzberg,O., Feizi,T. and Cambillau,C. (1994) *Nature Struct. Biol.*, 1, 863–870.

Chandra,N.R., Ramachandraiah,G., Bachhawat,K., Dam,T.K., Surolia,A. and Vijayan,M. (1999) *J. Mol. Biol.*, 285, 1157–1168.

Chelvanayagam,G., Heringa,J. and Argos,P. (1992) *J. Mol. Biol.*, 228, 220–242.

Connolly,M.L. (1993) *J. Mol. Graphics*, 11, 139–141.

Crennell,S., Garman,E., Laver,G., Vimr,E. and Taylor,G. (1994) *Structure*, 2, 535–544.

Divne,C., Ståhlberg,J., Reinikainen,T., Ruohonen,L., Pettersson,G., Knowles,J.K.C., Teeri,T.T. and Jones,T.A. (1994) *Science*, 265, 524–527.

Emsley,J., White,H.E., O'Hara,B.P., Oliva,G., Srinivasan,N., Tickle,I.J., Blundell,T.L., Pepys,M.B. and Wood,S.P. (1994) *Nature*, 367, 338–345.

Emsley,P., Potinou,C., Black,N., Fairweather,F., Charles,I.G., Wan,C., Hewitt,E. and Isaacs,W. (2000) *J. Biol. Chem.*, 275, 8889–8894.

Gibrat,J.-F., Madej,T. and Bryant,S.H. (1996) *Curr. Opin. Struct. Biol.*, 6, 377–385.

Grishkovskaya,I., Avvakumov,G.V., Sklenar,G., Dales, D., Hammond,G.L. and Muller,Y.A. (2000) *EMBO J.*, 19, 504–512.

Grochulski,P., Masson,L., Borisova,S., Pusztai-Carey,M., Schwartz,J-L., Brousseau,R. and Cygler,M. (1995) *J. Mol. Biol.*, 254, 447–464.

Hamelryck,T.W., Poortmans,F., Goossens,A., Angenon,G., Van Montagu,M., Wyns,L. and Loris,R. (1996) *J. Biol. Chem.*, 271, 32796–32802.

Hardman,K.D. and Ainsworth,C.F. (1972) *Biochemistry*, 11:4910–4919.

Havukainen,R., Torronen,A., Laitinen,T. and Rouvinen,J. (1996) *Biochemistry*, 35, 9617–9624.

Hohenester,E., Tiri,D., Talts,J.F. and Timpl,R. (1999) *Mol. Cell.*, 4, 783–792.

Holm,L. and Sander,C. (1995) *Trends Biochem. Sci.*, 20, 478–480.

Holm,L. and Sander,C. (1996) *Science*, 273, 595–602.

Keitel,T., Simon,O., Borriss,R. and Heinemann,U. (1993) *Proc. Natl Acad. Sci. USA*, 90, 5287–5291.

Kraulis,P.J. (1991) *J. Appl. Crystallogr.*, 24, 946–950.

Lis,H. and Sharon,N. (1998) *Chem. Rev.*, 98, 637–674.

Loris,R., Hamelryck,T., Bouckaert,J. and Wyns,L. (1998) *Biochim. Biophys. Acta*, 1383, 9–36.

Manoj,N., Srinivas,V.R., Surolia,A., Vijayan,M. and Suguna,K. (2000) *J. Mol. Biol.*, 302, 1129–1137.

Mourey,L., Pedelacq,J.D., Birck,C., Fabre,C., Rouge,P. and Samama,J.P. (1998) *J. Biol. Chem.*, 273, 12914–12922.

Murzin,A.G., Brenner,S.E., Hubbard,T. and Chothia,C. (1995) *J. Mol. Biol.*, 247, 536–540.

Orengo,C.A., Michie,A.D., Jones,S., Jones,D.T., Swindells,M.B. and Thornton,J.M. (1997) *Structure*, 5, 1093–1108.

Prabu,M.M., Sankaranarayanan,R., Puri,K.D., Sharma,V., Surolia,A., Vijayan,M. and Suguna,K. (1998) *J. Mol. Biol.*, 276, 787–796.

Prabu,M.M., Suguna,K. and Vijayan,M. (1999) *Proteins: Struct. Funct. Genet.*, 35,58–69

Romao,M.J., Kolln,I., Dias,J.M., Carvalho,A.L., Romero,A., Varela,P.F., Sanz,L., Topfer-Petersen,E. and Calvete,J.J. (1997) *J. Mol. Biol.*, 274, 650–660.

Romero,A., Romao,M.J., Varela,P.F., Kolln,I., Dias,J.M., Carvalho,A.L., Sanz,L., Topfer-Petersen,E. and Calvete,J.J. (1997) *Nature Struct Biol.*, 4, 783–788.

Rudenko,G., Nguyen,T., Chelliah,Y., Sudhof,T.C. and Deisenhofer,J. (1999) *Cell*, 99, 93–101.

Sharma,V. and Surolia,A. (1997) *J. Mol. Biol.*, 267, 433–445.

Shrive,A.K., Cheetham,G.M., Holden,D., Myles,D.A., Turnell,W.G., Volanakis,J.E., Pepys,M.B., Bloomer,A.C. and Greenhough,T.J. (1996) *Nature Struct Biol.*, 3, 346–354.

Srinivasan,N., Rufino,S.D., Pepys,M.B., Wood,S.P. and Blundell,T.L. (1996) *Chemtracts – Biochem. Mol. Biol.*, 6, 149–164.

Swaminathan,G.J., Leonidas,D.D., Savage,M.P., Ackerman,S.J. and Acharya,K.R. (1999) *Biochemistry*, 38, 13837–13843.

Umland,T.C., Wingert,L.M., Swaminathan,S., Furey,W.F., Schmidt,J.J. and Sax,M. (1997) *Nature Struct Biol.*, 4, 788–792.
Varela,P.R., Romero,A., Sanz,L., Romao,M.J., Topfer-Petersen,E. and Calvete,J.J. (1997) *J. Mol. Biol.*, 274, 635–649.
Vijayan,M. and Chandra,N. (1999) *Curr. Opin. Struct. Biol.*, 9, 707–714.
Wood,T.M. (1989) In Coughlan,M.P. (ed.). *Enzyme Systems for Lignocellulose Degradation*. Elsevier, London, pp. 19–35.

Received March 7, 2001; revised July 11, 2001; accepted July 31, 2001

Glycobiology vol. 14 no. 10 pp. 923–929, 2004
doi:10.1093/glycob/cwh114
Advance Access publication on June 16, 2004

Crystal structures of *Erythrina cristagalli* lectin with bound *N*-linked oligosaccharide and lactose

Kathryn Turton[3], Ramanathan Natesh[1,3], Nethaji Thiyagarajan[3], John A. Chaddock[4], and K. Ravi Acharya[2,3]

[3]Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, United Kingdom, and [4]The Health Protection Agency, Porton Down, Salisbury SP4 0JG, United Kingdom

Received on April 19, 2004; revised on June 1, 2004; accepted on June 11, 2004

Erythrina cristagalli **lectin (ECL) is a galactose-specific legume lectin. Although its biological function in the legume is unknown, ECL exhibits hemagglutinating activity *in vitro* and is mitogenic for T lymphocytes. In addition, it has been recently shown that ECL forms a novel conjugate when coupled to a catalytically active derivative of the type A neurotoxin from *Clostridium botulinum*, thus providing a therapeutic potential. ECL is biologically active as a dimer in which each protomer contains a functional carbohydrate-combining site. The crystal structure of native ECL was recently reported in complex with lactose and 2′-fucosyllactose. ECL protomers adopt the legume lectin fold but form non-canonical dimers via the handshake motif as was previously observed for *Erythrina corallodendron* lectin. Here we report the crystal structures of native and recombinant forms of the lectin in three new crystal forms, both unliganded and in complex with lactose. For the first time, the detailed structure of the glycosylated hexasaccharide for native ECL has been elucidated. The structure also shows that in the crystal lattice the glycosylation site and the carbohydrate binding site are involved in intermolecular contacts through water-mediated interactions.**

Key words: crystal structure/*Erythrina cristagalli* lectin/*Erythrina corallodendron* lectin/*N*-glycosylation/protein–carbohydrate interaction

Introduction

Erythrina cristagalli lectin (ECL) is a galactose-specific legume lectin. Although its function in the legume is unknown, *in vitro* ECL has been shown to have hemagglutinating activity and to be mitogenic for human T lymphocytes (Iglesias *et al.*, 1982). Recently, a conjugate comprising a catalytically active derivative of *Clostridium botulinum* neurotoxin A coupled to ECL has been used to selectively target nociceptive afferents (Duggan *et al.*, 2002). The specificity of ECL in retargeting the potent endopeptidase activity of botulinum neurotoxin A to nociceptive afferents *in vitro* points to the potential therapeutic use of this conjugate in the treatment of chronic pain.

ECL has been well studied in terms of carbohydrate binding. It interacts more strongly with fucosyllactose and fucosyllactosamine (Moreno *et al.*, 1997; Surolia *et al.*, 1996; Teneberg *et al.*, 1994) than with *N*-acetyllactosamine, lactose, *N*-acetylgalactosamine and galactose (Iglesias *et al.*, 1982). ECL differs subtly from the other *Erythrina* lectins in that it has a similar affinity for fucosyllactose and fucosyllactosamine, whereas other members of the family exhibit a preference for fucosyllactose (Moreno *et al.*, 1997). The crystal structure of native ECL (*n*ECL) has recently been determined in complex with lactose and fucosyllactose, providing insight into its altered carbohydrate specificity (Svensson *et al.*, 2002).

The *n*ECL protomer adopts a jelly-roll topology, in common with other legume lectins. Each protomer contains one Ca^{2+} and one Mn^{2+} ion, both of which are required for carbohydrate binding activity (Emmerich *et al.*, 1994). These ions are situated close to the carbohydrate-binding site (combining site) and help maintain the correct spatial orientation of combining site residues (Derewenda *et al.*, 1989). A *cis*-peptide bond between residues Ala88 and Asn89 holds the side chain of Asn89 in the correct orientation for carbohydrate binding (Svensson *et al.*, 2002).

*n*ECL has been shown to be glycosylated at Asn17 and Asn113, with partial occupancy at Asn113 and the nature of the bound heptasaccharide has been characterized (Ashford *et al.*, 1991). The major component of the carbohydrates bound to ECL contains two *N*-acetylglucosamine (GlcNAc) residues, one fucose, one xylose, and three mannose residues. Native ECL exists as a dimer in which two protomers associate back-to-back, forming a handshake motif. This noncanonical mode of dimerization was first observed in *Erythrina corallodendron* lectin (ECorL) (Shaanan *et al.*, 1991), which shares 96% sequence identity with *n*ECL. In ECorL, the ordered heptasaccharide bound to Asn17 is believed to prevent formation of the canonical dimer (Shaanan *et al.*, 1991) and it has been suggested that this may explain why *n*ECL dimers also adopt the handshake motif (Svensson *et al.*, 2002). However, the reported crystal structure of *n*ECL did not provide the structural details of the bound heptasaccharide.

Despite the importance of oligosaccharide interactions in glycoproteins, only little structural knowledge (using X-ray crystallography) has been gained over the years. This is mainly due to the inherent mobility and chemical heterogeneity of the oligosaccharides that prevent crystallization. However, in a few cases (such as ECorL) this has been achieved (Shaanan *et al.*, 1991). Here we report the

[1]Present address: ICGEB, Aruna Asaf Ali Marg, New Delhi 110 067, India
[2]To whom correspondence should be addressed; e-mail: k.r.acharya@bath.ac.uk

crystallization and structures of native and recombinant ECL (*n*ECL and *rec*ECL, respectively) in three new crystal forms. From the *n*ECL structure, we were able to glean the detailed picture of most of the bound oligosaccharide on Asn113 (a hexasaccharide portion of the heptasaccharide). The protomers of *n*ECL adopt the legume lectin fold and demonstrate structural equivalence of the recombinant form despite its lack of glycosylation. We examined the structural effects of glycosylation on quarternary structure by comparing the dimers formed by native and recombinant ECL. We confirm that dimers of both recombinant and native ECL adopt the handshake motif and suggest that structural factors other than the presence of glycosylation induce the protein to form a noncanonical dimer. Furthermore, we have studied the mode of lactose binding, which is similar to that observed for ECorL.

Results and discussion

nECL

*n*ECL cocrystallized with lactose in space group $P6_5$, with two molecules per crystallographic asymmetric unit and 67% of the crystal volume occupied by solvent. The final model (dimer) at 2.0 Å resolution (R_{cryst} 18.97%, R_{free} 20.92%) contains 477 amino acids, 2 calcium and manganese ions, 2 lactose molecules, 4 GlcNAc residues, 2 fucose residues, 1 xylose, 2 mannose residues, 4 HEPES molecules, and 456 water molecules (Table I). The estimated Luzzati coordinate error is 0.22 Å. Analysis of the Ramachandran plot revealed that the side chains of more than 99% of the amino acids have allowed or additional allowed conformations. The side chains of Tyr106 in each molecule adopt generously allowed conformations. The first three residues of the heptasaccharide bound to Asn113 were modeled into the electron density in one of the two protein molecules in the asymmetric unit, whereas electron density for six of the seven sugar residues was observed in the other.

recECL

*rec*ECL crystallized in triclinic form, with four molecules per unit cell (49% of the crystal volume occupied by the solvent). The final model at 2.13 Å resolution (R_{cryst} 19.1%, R_{free} 24.1%) contains 958 amino acids (molecule A, 1–239; molecule B, 1–239; molecule C, 1–240 and molecule D, 1–240), 4 calcium and manganese ions, 3 glycerol molecules, and 879 water molecules (Table I). The Luzzati coordinate error is 0.23 Å, and the average B factor for protein atoms is 24.52 Å^2. $|F_o| - |F_c|$ electron density in the combining site of three protomers was identified as glycerol, which had been included in the cryoprotectant solution used during data collection. A single residue, Tyr106, in only one of the four protomers is located in a generously allowed region of the Ramachandran plot. Possible alternative conformations were observed in the electron density map for the side chains of Ser120, but not in all of the molecules in the unit cell.

*rec*ECL was also cocrystallized with lactose in space group $P2_1$, with four molecules per asymmetric unit (49% solvent). The final model at 1.7 Å resolution (R_{cryst} 17.79%, R_{free} 20.31%) contains 959 amino acids, 4 lactose molecules, 4 calcium and manganese ions, and 1119 water molecules (Table I). The Luzzati coordinate error is 0.20 Å, and the average B factor for protein atoms is 19.99 Å^2. Five amino acid residues are located in generously allowed regions of the Ramachandran plot—Tyr106 from each protomer and Asp221 from one of the four protein molecules. Several side chains were observed to have potential alternative conformations (Met95, Asp161, Leu180, and His234) but not in all of the protomers.

Table I. Crystallographic data processing and refinement statistics

	*n*ECL	*rec*ECL	*rec*ECL
Data processing			
Space group	$P6_5$	P1	$P2_1$
Cell dimensions (Å)	a = 134.02	a = 55.28	a = 54.90
	b = 134.02	b = 55.37	b = 167.23
	c = 81.64	c = 86.93	c = 55.13
	α = β = 90°	α = 86.23°	α = γ = 90°
	γ = 120°	β = 75.37°	β = 97.09°
		γ = 82.13°	
Resolution (Å)	2.00	2.13	1.70
Reflections measured	414,754	399,699	1,021,770
Unique reflections	56,451	55,200	108,695
Completeness (%)	99.7	93.1	94.1
In the outermost shell	99.9	86.1	77.3
I/(σI)	18.44	10.93	28.05
In the outermost shell	6.22	5.29	5.26
R_{symm}	0.070	0.077	0.058
Refinement			
R_{cryst} (%)	18.97	19.10	17.79
R_{free} (%)	20.92	24.10	20.31
Average B factor (Å^2)	28.84	24.52	19.99
Wilson	23.1	25.60	19.20
RMSD from ideality			
Bonds (Å)	0.005	0.009	0.006
Bond angles (deg)	1.44	1.47	1.49
Dihedrals (deg)	25.32	26.15	26.10
Impropers (deg)	0.78	0.91	0.78

$R_{symm} = \Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I\rangle(hkl)|/\Sigma_{hkl}\Sigma_i I_i(hkl)$ where $\langle I\rangle$ is the averaged intensity of the i observations of reflection hkl.
$R_{cryst} = \Sigma||F_O| - |F_C||/\Sigma|F_O|$ where F_O and F_C are the observed and calculated structure factor amplitudes, respectively.
R_{free} is equal to R_{cryst} for a randomly selected 5% subset of reflections not used in the refinement.

Overall structure

Protomers of ECL adopt the conserved jelly-roll fold that is characteristic of legume lectins (Figure 1). This structural motif comprises a six-stranded back β-sheet, a curved seven-stranded front β-sheet, a short five-membered β-sheet, and a set of loops connecting the three sheets. Both the six- and seven-stranded β-sheets are entirely

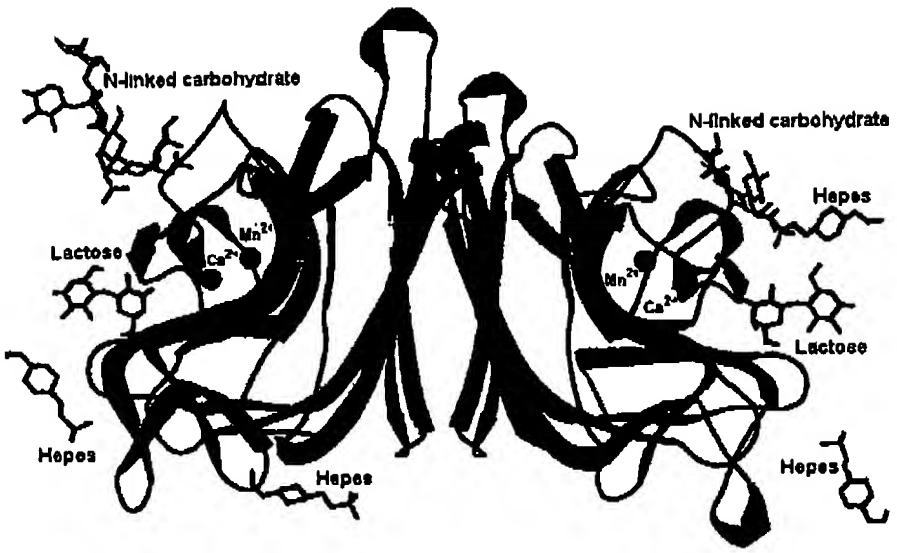


Fig. 1. The structure of ECL dimer. Protomers of ECL associate together back-to-back to form noncanonical dimers via the handshake motif as was first observed for ECorL. The dimer structure is stabilized by two hydrogen bonds and a series of contacts between side chains in four strands of the flat, six-stranded β-sheet. The protomers are tilted with respect to one another such that the N- and C-termini play no part in the dimer interface. Each protomer adopts the conserved jelly-roll fold characteristic of legume lectins. The *N*-linked carbohydrate (on Asn113), lactose bound at the combining site and HEPES molecules (from the crystallization medium) are shown. The manganese and calcium ions bound in the vicinity of the combined site are shown as small spheres.

antiparallel. The N- and C-termini associate together, forming the first two strands of the six-stranded β-sheet.

The crystal structures of *n*ECL and *rec*ECL at 2.0 Å and 2.13 Å resolution, respectively, superimpose with an overall root mean squared deviation of 0.26 Å (Cα atoms). Comparison of the two structural models indicated that there are no significant differences between the crystal structures of *n*ECL and *rec*ECL, thus confirming that *rec*ECL adopts native-like structure. The main difference between the native and recombinant forms is the absence of *N*-linked carbohydrate in *rec*ECL.

Each ECL protomer contains one Ca^{2+} and one Mn^{2+} ion, both located close to the carbohydrate combining site. The metal ions are approximately 4.2 Å apart, and each coordinate two water molecules and make contacts with four amino acids in the vicinity of the combining site. One of these structural waters also stabilizes the conserved *cis*-peptide bond (Ala88–Asp89) that correctly orients the side chain of Asp89 for carbohydrate binding. Mn^{2+} makes contacts with the side chains of Glu127, Asp129, Asp136, and His142, and Ca^{2+} makes contacts with Asp129, Phe131, Asn133, and Asp136 (Table II). In addition to binding the calcium ion, the side chains of Phe131 and Asn133 are also implicated in lactose binding (see following discussion).

ECL exists as a biologically active dimer and has not been isolated in monomeric form. Both *n*ECL and *rec*ECL form noncanonical back-to-back dimers via the handshake motif in which the two protomers are tilted with respect to each other (Figure 1). This mode of dimerization has previously been reported for ECorL and the basic winged bean lectin (Prabu *et al.*, 1998; Shaanan *et al.*, 1991). The dimer interface exists between the six-stranded sheets of each protomer and is stabilized by two hydrogen bonds between the side chain of Lys171 in one protomer and Thr193 in the other and a number of van der Waals contacts between residues in four strands of the six-stranded β-sheet—the strands formed by the N- and C-termini do not take part in stabilizing the quarternary structure. Residues Arg73, Glu79, Gln80, Pro81, Tyr82, Thr83, Arg84, Lys116, Gln117, Asp118, Asn119, Asn148, Asp161, Asn162, Gln164, Lys171, Ile191, Thr193, Gln202, Val203, and Asp221 mediate these contacts at the dimer interface.

Binding of N-linked glycosylated saccharide

The *N*-linked oligosaccharide bound to *n*ECL is covalently bound to Asn113, which is part of the loop structure located between strands β5 and β6. This residue is the only possible site of attachment for *N*-linked glycosylation in the 241-amino-acid sequence of ECL. In one of the two protein molecules (molecule B) in the asymmetric unit, there was enough electron density to model six of the seven sugar residues bound to Asn113 (Figure 2). The modeled hexasaccharide has the profile: αD-Man-(1 → 3)-[β-D-Xyl-(1 → 2)]-β-D-Man-β-D-GlcNAc-(1 → 4)-[α-L-Fuc (1 → 3)]-D-GlcNAc and does not make contacts with any other parts of the lectin protomer to which it is bound, although it might communicate with the lactose moiety bound in the combining site of a symmetry-related molecule through water mediated interactions.

Influence of glycosylation on the quarternary structure

ECL shares 96% sequence identity with ECorL, which was thought to dimerize in a noncanonical fashion because of

Table II. Contacts between ECL and metal ions

Metal ion	Amino acid side chain	Distance (Å)
Ca^{2-}	Asp136 OD2	2.40
	Asp129 OD2	2.53
	Asp 129 OD1	2.60
	Asn133 OD1	2.45
	Water	2.54
	Water	2.55
	Asp129 CG	2.9
	Asp129 OD1	2.6
	Asp129 OD2	2.5
	Phe131 O	2.5
	Asn133 OD1	2.5
Mn^{2+}	His142 NE2	2.39
	Glu127 OE2	2.22
	Asp129 OD2	2.23
	Asp136 OD1	2.18
	Water	2.33
	Water	3.96
	Water	2.21
	Glu127 OE2	2.2
	Asp129 CG	3.2
	Asp129 OD2	2.2
	Asp136 CG	3.1
	Asp136 OD1	2.2
	His142 NE2	2.4
	His142 CE1	3.2

Contacts formed between ECL and the calcium and manganese ions bound close to the carbohydrate combining site. Each of the metal ions is bound close to the ECL combining site through interactions with four amino acid side chains and two water molecules, one of which also stabilizes the conserved *cis*-peptide bond.

the glycosylation attached to Asn17. Based on their high sequence identity, one would expect ECL to dimerize in the same way as ECorL, but the form of ECL studied in this investigation lacks the *N*-linked glycosylation site at position 17 (which is located at the opposite end of the molecule compared with Asn113 in ECL). This raises an interesting issue regarding its mode of dimerization: If there is no heptasaccharide bound to residue 17, what might force ECL to form a noncanonical dimer? Several studies have been undertaken to rationalize the various oligomerization states of legume lectins in terms of their amino acid sequences, shape complementarity of protomers, interaction energy between protomers, and hydrophobic surface area buried on oligomerisation (Elgavish and Shaanan, 2001; Manoj and Suguna, 2001; Prabu *et al.*, 1999; Srinivas *et al.*, 2001). The results of these studies indicate that the observed modes of oligomerization are energetically more favorable than any alternative quarternary structures.

We confirm that both native and recombinant forms of ECL associate into dimers back-to-back via the handshake motif. Because *rec*ECL is unglycosylated, this demonstrates that the presence of *N*-linked glycosylation does not influence the mode of dimerization in ECL and suggests that factors intrinsic to the primary structure of the lectin dictate its quarternary structure. Examination of legume lectin sequences, focusing on the regions forming interfaces between protomers and dimers, has identified amino acid residues that might influence the mode of oligomerization (Manoj and Suguna, 2001). Extrapolation of these results to ECL reveals that the primary structure of this lectin contains features (Glu2, Glu12, Lys55, Arg73, and Lys171) that indicate it could be expected to form ECorL-type dimers. Lectins that do not form canonical dimers have an acidic residue at the position corresponding to Lys55 in ECL and a charged residue equivalent to Glu12 that comes into close contact with a charged residue (Glu2) that would form unfavorable interactions if a concanavalin A–type dimer were formed (Manoj and Suguna, 2001). Arginine and lysine residues at positions equivalent to 73 and 171 in ECL are conserved among lectins forming ECorL-type dimers.

Lactose binding

The combining site is located in a shallow cleft on the surface of ECL and accommodates the galactose moiety of bound carbohydrates. The crystal structures of the *n*ECL-lactose and *rec*ECL-lactose complexes superimpose with an overall root mean squared deviation of 0.27 Å. The spatial arrangement of residues in the combining sites of both forms of the lectin are identical, confirming that *rec*ECL binds lactose in the same way as *n*ECL. Both *n*ECL and *rec*ECL bind lactose through a set of structural water molecules. These water molecules mediate indirect hydrogen bonds between Gly107, Asn133, Ala218, and Gln219 and the O2, O3, and O6 of galactose and O2 of glucose. Lactose also makes contacts with more water molecules in the combining site (Figure 3). Hydrophobic stacking interactions were observed between the aromatic ring of Phe131, the galactose ring of the bound lactose and the side chain of Tyr106, forming a sandwich with the galactose ring between the aromatic side chains. The side chain of Tyr106 adopts a generously allowed configuration in the lactose-bound structures, as determined by the Ramachandran plot. In the unliganded *rec*ECL structure, Tyr106 in only one of the four protein molecules was located in an additional allowed region of the plot. Thus it appears that the conformation of Tyr106 is affected by carbohydrate binding.

The mode of lactose binding in ECL is similar to that observed for ECorL (Shaanan *et al.*, 1991), although there are subtle differences in their carbohydrate specificities (Moreno *et al.*, 1997; Teneberg *et al.*, 1994). The altered carbohydrate specificity of ECL compared to ECorL is postulated to be due to differences in their amino acid sequences at positions 111 and 125 (Svensson *et al.*, 2002)—substitution of residues at these positions causes rotation of the side chain of Val92, which is thought to induce structural changes in the combining site. Analysis of the crystal structure of *n*ECL revealed that although there are no contacts between these three residues, Val92 makes contacts with Val126, which is located close to the combining site. It is possible that substitutions causing

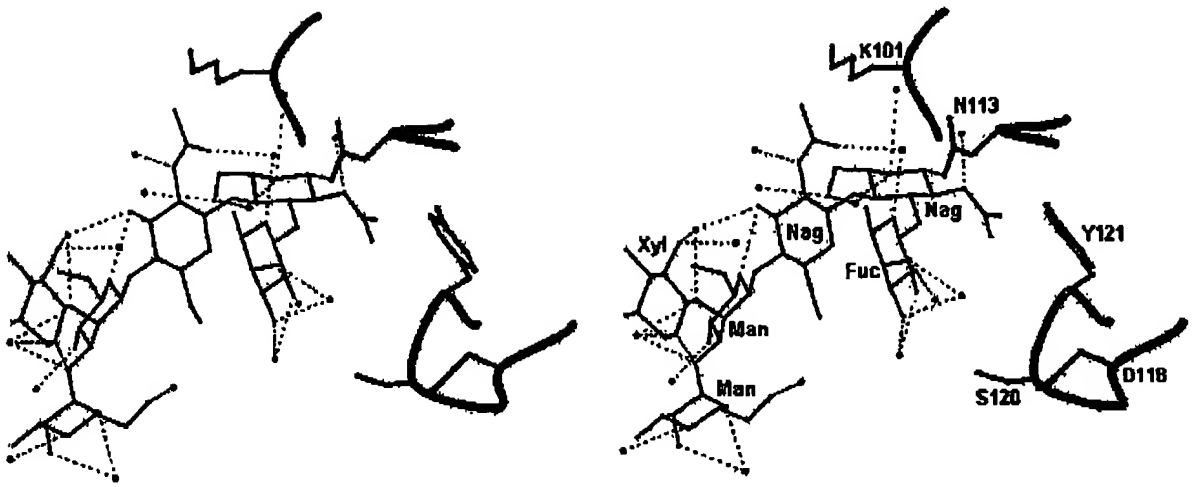


Fig. 2. Structure of bound *N*-linked hexasaccharide to *n*ECL. Details of the surrounding residues for the hexasaccharide at the glycosylation site (Asn113). In the crystal lattice, the N-linked carbohydrate makes indirect interaction with the combining site through water-mediated interaction.

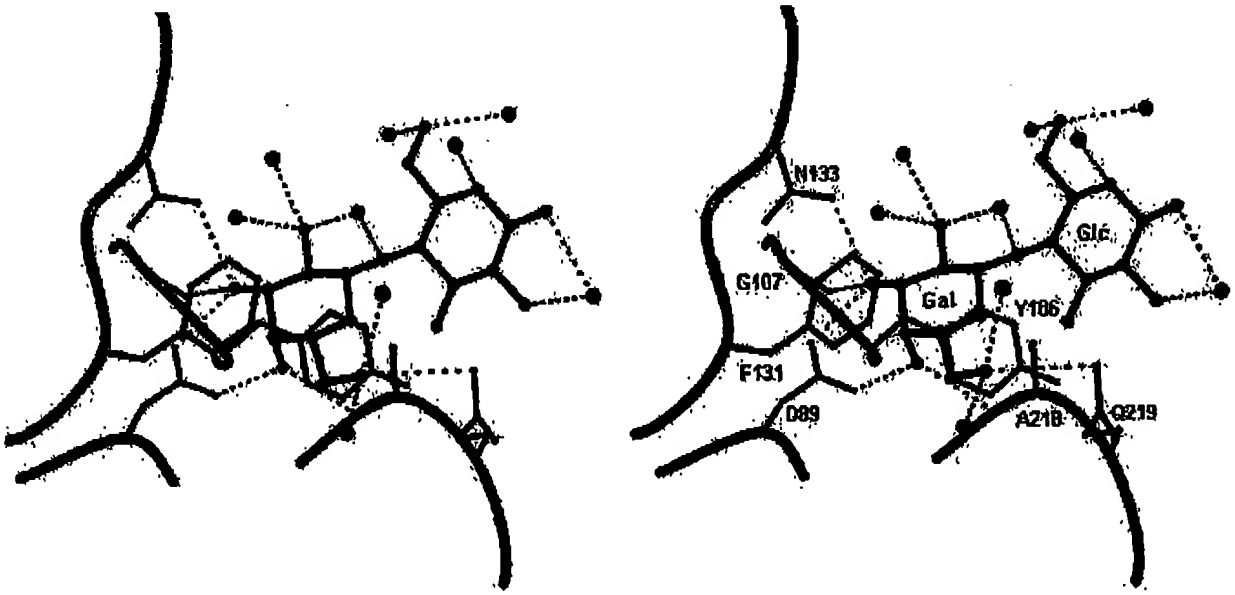


Fig. 3. Binding of lactose to ECL. Hydrophobic interactions were observed between the side chains of Tyr106 and Phe131 and the galactose ring of the lactose molecule. This sandwiches the sugar ring between the aromatic side chains of the two amino acids. Lactose makes a number of contacts with side chains of combining site residues, but only the galactose moiety is accommodated in the shallow cleft of the combining site. There are no direct hydrogen bonds between the lectin and disaccharide; instead, interactions are mediated by a set of structural water molecules. Key residues involved in carbohydrate binding are Leu86, Asp89, Gly107, Asn133, Ala218, and Gln219.

movement of Val92 might indirectly affect carbohydrate binding. However, the real effect of Val92 on carbohydrate binding remains to be fully analyzed.

The structures of *rec*ECL unliganded and in complex with lactose (at 2.13 Å and 1.70 Å resolution, respectively) superimpose with an overall root mean squared deviation of 0.21 Å. A direct comparison of side chain positions in the combining sites of the two structures revealed that there are no structural rearrangements on lactose binding. This indicates that the amino acid residues involved in carbohydrate recognition are optimally oriented in the ECL protomer.

Comparison with previously reported structure of nECL in the presence of lactose

As already mentioned, the crystal structures of *n*ECL in complex with lactose and 2′-fucosyllactose were previously

reported (Svensson *et al.*, 2002). Comparison of the amino acid sequence of *n*ECL used in this investigation with that reported for native ECL (hereafter referred to as *Sven*ECL) revealed 10 differences: Asn16Asp, Asp17Asn, Ile25Leu, Ile59Met, Met62Ser, Ser175Pro, Leu180His, Ala181Val, Glu206Asp, and His234Gln (where the first residue is that in *n*ECL and the second is that in *Sven*ECL). The differences at positions 25, 181, and 206 represent conservative substitutions. The residues at positions 16 and 17 appear to be swapped in *n*ECL compared with *Sven*ECL, but because the side chains are of the same size and shape, these amino acids cannot be distinguished from one another on the basis of their electron density. The sequence of *rec*ECL was determined by DNA sequencing of multiple clones from two independent sources of *E. cristagalli* (Stancombe *et al.*, 2002) and analysis of the electron density maps confirmed this sequence at all 10 locations where it differs from *Sven*ECL (obtained by mass spectrometric peptide mapping, tandem mass spectrometry, and X-ray structure; Svensson *et al.*, 2002). However, it is important to note that none of the sequence differences affect amino acids that are involved in carbohydrate binding or stabilizing the dimer interface. The main difference is the absence of a potential *N*-linked glycosylation site at residue 17 from *n*ECL. The electron density maps for *n*ECL were carefully inspected to ensure that there was no evidence of residue 17 being asparagine instead of aspartic acid, but no density was observed that might represent bound carbohydrate. Therefore we are confident that the sequence of *n*ECL matches that of *rec*ECL.

Conclusions

In this investigation, *n*ECL was crystallized in a new crystal form in complex with lactose and *rec*ECL was crystallized for the first time in two different crystal forms, both unliganded and in complex with lactose. We confirm that the tertiary structure of both forms of ECL is homologous to the known structures of other legume lectins, with protomers adopting the jelly-roll (legume lectin) fold. We have confirmed that *n*ECL protomers associate back-to-back via the handshake motif, and we have shown that *rec*ECL protomers dimerize in the same way. Following comparison of native and recombinant forms of the lectin, it can be concluded that the presence of bound oligosaccharide does not influence the tertiary or quarternary structure of ECL. Furthermore, comparison of the structures of *n*ECL and *rec*ECL in complex with lactose confirms that the presence of *N*-linked glycosylation on *n*ECL has no major effect on the structure of the combining site or carbohydrate binding. Thus we confirm that *rec*ECL is native-like in terms of both its structure and biological activity.

Materials and methods

Protein purification and crystallization

*n*ECL was purchased from Sigma (Dorset, UK). *rec*ECL was expressed and purified as previously reported (Stancombe *et al.*, 2003). Briefly, *E. cristagalli* seeds (Sandeman Seeds, Oxford Botanical Gardens, UK) were germinated and genomic DNA, obtained from leaf material, was amplified by polymerase chain reaction and cloned into vector pMTL1015. For expression, this clone was transformed into *Escherichia coli* BL21 (DE3) cells and cultured at 8-L fermentation scale at 37°C until static growth ($OD_{600} = 30$). The protein was solubilized from inclusion bodies and subsequently refolded and purified on an immobilized lactose matrix. After extensive washing, *rec*ECL was eluted by the addition of 0.3 M lactose.

Crystallization was achieved using the vapor diffusion method (hanging drops) at 16°C, and crystals were observed within 3–4 weeks. For crystallization of *n*ECL in space group $P6_5$, drops made up of 2 ml protein solution, 2 ml mother liquor (70% 2-methyl-2,4-pentanediol and 0.1 M HEPES, pH 7.0) and 0.4 ml 100 mM lactose solution and were equilibrated over sealed wells containing 800 ml mother liquor. *rec*ECL crystallized in space group P1 from drops containing 2 ml protein and 2 ml mother liquor (17% polyethylene glycol [PEG] 3350, 0.3 M sodium chloride, and 0.02 M imidazole). *rec*ECL was also cocrystallized with lactose using mother liquor made up of 20% PEG 3350 and 0.2 M imidazole.

X-ray data collection and structure determination

X-ray diffraction data were collected at 100 K using the Synchrotron radiation Source at Daresbury, UK, on station PX14.1 (wavelength 1.483 Å). Raw data images for *n*ECL were indexed and integrated using DENZO (Otwinowski and Minor, 1997) and then scaled with SCALEPACK (Otwinowski and Minor 1997). X-ray images collected for the *rec*ECL crystals were processed and scaled using HKL2000 (Otwinowski and Minor, 1997) (Table 1).

Initial phases were obtained by the molecular replacement method using the program AMoRe (Navaza, 1994). For *n*ECL, the crystal structure of ECorL at 1.95 Å resolution (PDB code 1AX1) (Elgavish and Shaanan, 1998) was used as an initial model with the nine nonconserved residues mutated to alanine, and all waters, ligands, and metal ions removed. Refinement was performed using the CNS suite of programs (Brunger *et al.*, 1998). After one round of refinement, the crystallographic R factor (R_{cryst}) dropped to 29.59% (R_{free} 32.38%, based on 5% of reflections omitted from the refinement). Calculated phases from the refined structure were used to determine $|F_o| - |F_c|$ and $2|F_o| - |F_c|$ electron density maps. Careful examination of the $|F_o| - |F_c|$ map allowed mutation of the nonconserved residues back to their native side chains and the addition of calcium and manganese ions to the model. HEPES and lactose were built into electron density of each protomer, with parameter and topology files from the HIC-Up server (Kleywegt and Jones, 1998). Similarly glycosylated sugars (three in molecule A and six in molecule B, part of the heptasaccharide bound to Asn113) were modeled into the $|F_o| - |F_c|$ electron density map, riding on the Asn113 residue.

Repeated rounds of refinement and model building were carried out to improve the model. After several rounds of refinement, water molecules were added to the structure if there were peaks in the $|F_o| - |F_c|$ electron density maps with heights greater than 3Σ at hydrogen bond forming distances from the appropriate atoms. $2|F_o| - |F_c|$ maps were also used to check the consistency in peaks. Water

molecules with a temperature factor of >60 Å^2 were excluded from subsequent refinement steps.

Subsequent solution of *rec*ECL structures was achieved using the refined coordinates of *n*ECL (at 2.0 Å resolution) as a search model. Model building and refinement procedures were undertaken as described for *n*ECL. For *rec*ECL in space group P1, apart from the lactose molecule, glycerol was modeled into the carbohydrate-combining site.

The program PROCHECK (Laskowski *et al.*, 1993) was used to assess the quality of each structure after the final round of refinement. Analysis of the Ramachandran plot for each structure revealed that over 99% of residues were located in allowed regions of the plot. The refinement statistics for all structures are listed in Table I.

Acknowledgments

We thank Frances Alexander for supply of purified *rec*ECL; the staff at the Synchrotron Radiation Source, Daresbury; and Gayatri Chavali and Shalini Iyer for help during X-ray data collection. This work was supported by postgraduate studentships to K.T. and N.T. through a joint agreement between the University of Bath and Health Protection Agency, Salisbury, United Kingdom. The atomic coordinates for *n*ECL ($P6_5$), *rec*ECL (P1), and *rec*ECL ($P2_1$) with codes 1UZY, 1UZZ, and 1V00, respectively, have been deposited with the RCSB Protein Data Bank.

Abbreviations

ECL, *Erythrina cristagalli* lectin; ECorL, *Erythrina corallodendron* lectin.

References

Ashford, D.A., Dwek, R.A., Rademacher, T.W., Lis, H., and Sharon, N. (1991) The glycosylation of glycoprotein lectins. Intra- and inter-genus variation in *N*-linked oligosaccharide expression. *Carbohydrate Res.*, 213, 215–227.

Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski J., Nilges, M., Pannu, N.S., and others. (1998) Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr.*, D54, 905–921.

Derewenda, Z., Yariv, J., Helliwell, J.R., Kalb, A.J., Dodson, E.J., Papiz, M.Z., Wan, T., and Campbell, J. (1989) The structure of the saccharide-binding site of concanavalin A. *EMBO. J.*, 8, 2189–2193.

Duggan, M.J., Quinn, C.P., Chaddock, J.A., Purkiss, J.R., Alexander, F.C.G., Doward, S., Fooks, S.J., Friis, L.M., Hall, Y.H.J., Kirby, E.R., and others. (2002) Inhibition of release of neurotransmitters from rat dorsal root ganglia by a novel conjugate of a *Clostridium botulinum* toxin A endopeptidase fragment and *Erythrina cristagalli* lectin. *J. Biol. Chem.*, 277, 34846–34852.

Elgavish, S. and Shaanan, B. (1998) Structures of the *Erythrina corallodendron* lectin and of its complexes with mono- and disaccharides. *J. Mol. Biol.*, 277, 917–932.

Elgavish, S. and Shaanan, B. (2001) Chemical characteristics of dimer interfaces in the legume lectin family. *Prot. Sci.*, 10, 753–761.

Emmerich, C., Helliwell, J.R., Redshaw, M., Naismith, J.H., Harrop, S.J., Raftery, J., Kalb, A.J., Yariv, J., Dauter, Z., and Wilson, K.S. (1994) High-resolution structures of single-metal-substituted concanavalin A: the Co,Ca-protein at 1.6Å and the Ni,Ca-protein at 2.0Å. *Acta Crystallogr.*, D50, 749–756.

Iglesias, J.L., Lis, H., and Sharon, N. (1982) Purification and properties of a D-galactose/N-acetyl-D-galactosamine-specific lectin from *Erythrina cristagalli*. *Eur. J. Biochem.*, 123, 247–252.

Kleywegt, G. and Jones, T.A. (1998) Databases in protein crystallography. *Acta Crystallogr.*, D54, 1119–1131.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., and Thornton, J.M. (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Cryst.*, 26, 283–291.

Manoj, N. and Suguna, K. (2001) Signature of quarternary structure in the sequences of legume lectins. *Protein Eng.*, 14, 735–745.

Moreno, E., Teneberg, S., Adar, R., Sharon, N., Karlsson, K.A., and Angstrom, J. (1997) Redefinition of the carbohydrate specificity of *Erythrina corallodendron* lectin based on solid-phase binding assays and molecular modeling of native and recombinant forms obtained by site-directed mutagenesis. *Biochemistry*, 36, 4429–4437.

Navaza,, J. (1994) AMoRe: an automated package for molecular replacement. *Acta Crystallogr.*, A50, 157–163.

Otwinowski, Z. and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, 276, 307–326.

Prabu, M.M., Sankaranarayanan, R., Puri, K.D., Sharma, V., Surolia, A., Vijayan, M., and Suguna, K. (1998) Carbohydrate specificity and quarternary association in basic winged bean lectin: X-ray analysis of the lectin at 2.5 Å resolution. *J. Mol. Biol.*, 276, 787–796.

Prabu, M.M., Suguna, K., and Vijayan, M. (1999) Variability in quaternary association of proteins with the same tertiary fold: A case study and rationalization involving legume lectins. *Prot. Struct. Funct. Genet.*, 35, 58–69.

Shaanan, B., Lis, H., and Sharon, N. (1991) Structure of a legume lectin with an ordered *N*-linked carbohydrate in complex with lactose. *Science*, 254, 862–866.

Srinivas, V.R., Reddy, G.B., Ahmad, N., Swaminathan, C.P., Mitra, N., and Surolia, A. (2001) Legume lectin family, the "natural mutants of the quarternary state", provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys. Acta*, 1527, 102–111.

Stancombe, P.R., Alexander, F.C.G., Ling, R., Matheson, M., Shone, C.C., and Chaddock, J.A. (2003) Isolation of the gene and large scale expression and purification of recombinant *Erythrina cristagalli* lectin. *Protein Exp. Purif.*, 30, 283–292.

Surolia, A., Sharon, N., and Schwarz, F.P. (1996) Thermodynamics of monosaccharide and disaccharide binding to *Erythrina corallodendron* lectin. *J. Biol. Chem.*, 271, 17697–17703.

Svensson, C., Teneberg, S., Nilsson, C.L., Kjellberg, A., Schwarz, F.P., Sharon, N., and Krengel, U. (2002) High-resolution crystal structures of *Erythrina cristagalli* lectin in complex with lactose and 2′-alpha-L-fucosyllactose and correlation with thermodynamic binding data. *J. Mol. Biol.*, 32, 69–83.

Teneberg, S., Angstrom, J., Jovall, P.A., and Karlsson, K.A. (1994) Characterization of binding of Gal beta 4GlcNAc-specific lectins from *Erythrina cristagalli* and *Erythrina corallodendron* to glycosphingolipids. Detection, isolation, and characterization of a novel glycosphinglipid of bovine buttermilk. *J. Biol. Chem.*, 269, 8554–8563.

Volume 257, number 1, 81–85 FEB 07725 October 1989

The amino acid sequence of *Erythrina corallodendron* lectin and its homology with other legume lectins*

Rivka Adar, Michael Richardson+, Halina Lis and Nathan Sharon

Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel and +*Department of Biological Sciences, University of Durham, Science Laboratories, South Road, Durham DH1 3LE, England*

Received 1 September 1989

The primary sequence of *Erythrina corallodendron* lectin was deduced from analysis of the peptides derived from the lectin by digestion with trypsin, chymotrypsin, *Staphylococcus aureus* V8 protease, elastase and lysylendopeptidase-C, and of fragments generated by cleavage of the lectin with dilute formic acid in 6 M guanidine hydrochloride. Purification of the individual peptides was achieved by gel filtration, followed by reverse phase HPLC. The glycosylation site (Asn^{17}-Leu^{18}-Thr^{19}) was deduced from analysis of the glycopeptide isolated from a pronase digest of the lectin before and after deglycosylation of the glycopeptide with endoglycosidase F. Comparison of the sequence of 244 residues thus obtained with those of 9 other legume lectins revealed extensive homologies, including 39 invariant positions and 60 partial identities. These data provide further evidence for the conservation of the lectin gene in leguminous plants.

Primary sequence; Sequence homology; Lectin, legume; Glycosylation site

1. INTRODUCTION

Erythrina is a family of deciduous leguminous trees and shrubs widely spread in the tropics and subtropics. Since 1980, lectins from some 20 species of this family have been isolated in different laboratories, a dozen of which by us [1,2]. All *Erythrina* lectins studied are specific for galactose and *N*-acetylgalactosamine and show pronounced preference for *N*-acetyllactosamine. They are glycoproteins (3–10% carbohydrate) of molecular masses in the range of 56000–68000 Da and are composed of two identical or nearly identical subunits. Whenever examined, the carbohydrate of the lectins was found to consist of glucosamine, mannose, L-fucose and xylose, predominantly in the form of the asparagine-linked heptasaccharide Manα3(Manα6)(Xylβ2)Manβ4GlcNAcβ4 (Fucα3)GlcNAc [3]. The N-terminal amino acid sequences, determined on 10 *Erythrina* lectins for up to 15 amino acids have been nearly identical [2].

Correspondence address: N. Sharon, Department of Biophysics, The Weizmann Institute of Science, Rehovot 76100, Israel

Abbreviations: TPCK, L-1-tosylamide-2-phenylethyl-chloromethyl ketone; SDS-PAGE, SDS-polyacrylamide gel electrophoresis; HPLC, high-performance liquid chromatography; ConA, concanavalin A; DBL, *Dolichos biflorus* lectin; ECorL, *Erythrina corallodendron* lectin; LCL, *Lens culinaris* (lentil) lectin; PHA-L, leukoagglutinin from *Phaseolus vulgaris*; PNA, peanut agglutinin; PSL, *Pisum sativum* (pea) lectin; SBA, soybean agglutinin; SL, sainfoin (*Onobrychis viciifolia* lectin)

* Dedicated to the memory of Professor Edgar Lederer (1908–1988)

We have now established the nearly complete sequence (the first 244 amino acids) of *Erythrina corallodendron* lectin (ECorL), as well as the glycosylation site. Comparison of this sequence with those of 9 other legume lectins reveals, as predicted [4], extensive homologies.

2. MATERIALS AND METHODS

2.1. *Preparation of ECorL*

The lectin extracted from the seeds of *Erythrina corallodendron* was purified by fractional precipitation with ammonium sulfate and affinity chromatography on a column of lactose coupled to divinylsulphone-activated Sepharose [2].

2.2. *Enzyme digestion*

The following digestions were performed, all at an enzyme/substrate ratio of 1:50. (i) ECorL (2 mg/ml) was heat-denatured at pH 2 [5] and digested with lysylendopeptidase-C (Calbiochem) in phosphate buffer, pH 7.8, for 6 h at 37°C [6]. (ii) Samples of ECorL, 10–15 mg in 1 ml of 0.2 M *N*-ethylmorpholine HCl buffer, pH 8.5, were digested separately with trypsin (TPCK-treated, Sigma) and chymotrypsin (α-chymotrypsin, type VIII, TPCK-treated, Sigma), for 5–10 h at 37°C. (iii) ECorL, 7 mg/ml of 0.1 M sodium bicarbonate, pH 8.1, was digested with *Staphylococcus aureus* V8 protease (Miles) at 37°C. Half of the sample was removed after 24 h, while the remainder was kept for 5 days at room temperature. (iv) The lectin, 10 mg in 2 ml of 0.2 M Tris-HCl, pH 8.8, was denatured by boiling for 10 min and digested with elastase (Sigma) for 75 min at 37°C.

2.3. *Chemical cleavage*

ECorL, 10 mg/ml, was cleaved with 75% formic acid in 6 M guanidine hydrochloride for 72 h at 37°C.

2.4. *Peptide separation and purification*

For the separation of the peptides in the lysylendopeptidase-C digest, aliquots (100 μg/well) were loaded onto 8–25% gradient

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50 © 1989 Federation of European Biochemical Societies

polyacrylamide slab gels containing 0.1% SDS and electrophoresed as described [7]. The separated peptides were electrophoretically transferred to a polyvinylidene difluoride membrane (PVDF, Millipore). Strips of the membranes were stained with Coomassie blue and bands corresponding to the peptides were cut out from the membrane [8].

Peptides in the tryptic, chymotryptic and elastase digests of ECorL were fractionated on columns (1 × 200 cm) of Biogel P-6 in 0.1 M ammonium bicarbonate pH 8.1 as in [9], while fragments resulting from acid hydrolysis were fractionated on a column (1 × 200 cm) of Biogel P-30 in 70% formic acid [10]. Elution was followed by measuring the absorbance at 230 nm. Fractions corresponding to the peaks were collected, lyophilized and dissolved in 0.1% trifluoroacetic acid for further purification by reverse-phase HPLC. The peptides in the *S. aureus* V8 digest were purified directly by reverse-phase HPLC.

HPLC was carried out on a Vydac analytical reverse-phase column (25 cm × 4.6 mm; 218TP54 HP Genenchem, San Francisco, CA) in a Waters 600E multisolvent delivery system using variable gradients of 0–50% acetonitrile (HPLC grade, Merck) in 0.1% trifluoroacetic acid [9,11]. Peptides were detected by measuring the absorbance at 214 nm, and collected manually directly from the UV monitor.

2.5. *Preparation of glycopeptide*

ECorL (500 mg, containing 3% neutral sugar) was denatured and digested with 10 mg of pronase (Calbiochem) as described [5]. The digest was lyophilized, the dry material dissolved in 7 ml of 0.01 M acetic acid and insoluble material removed by centrifugation. The supernatant was applied in two portions to a Sephadex G-50 column (2 × 160 cm) using 0.01 M acetic acid as a solvent. Fractions were collected, monitored by absorbance at 230 nm, and examined for their neutral sugar content by the phenol-sulfuric method [12], using mannose as standard. Sugar-containing fractions were pooled, lyophilized and rechromatographed on the same column, under the same conditions as described above.

2.6. *Deglycosylation*

An aliquot of the rechromatographed material (575 μg of mannose) was lyophilized, dissolved in 100 μl of 20 mM potassium phosphate buffer, pH 6.5, 25 mM EDTA and treated with 1 U of endoglycosidase F (Boehringer-Mannheim) for 20 h at 37°C. The enzyme-treated, as well as a sample of the untreated material, were analysed by reverse-phase HPLC using a gradient of 0–7% acetonitrile in 0.1% trifluoroacetic acid.

2.7. *Sequence determination*

Peptides derived from the various digests were subjected to microsequence analysis using the 4-*N*,*N*-dimethylaminoazobenzene-4'-isothiocyanate (DABITC)/phenylisothiocyanate (PITC) double-coupling method, followed by thin-layer chromatographic identification of the amino acid derivatives liberated [13,14]. The glycopeptide purified by HPLC, its endoglycosidase digestion product and the PVDF membranes containing peptides from the lysylendopeptidase-C digest were sequenced on a gas phase Applied Biosystem automatic sequencer model 470 A.

3. RESULTS AND DISCUSSION

3.1. *Sequence determination*

The sequence of 244 amino acids, together with the details of the overlapping peptides and fragments from which it was deduced is shown in fig.1. Digestion with trypsin and *S. aureus* V8 protease yielded peptides from which most of the sequence of ECorL could readily be established. Digestions with chymotrypsin, lysylendopeptidase-C and elastase gave good overlaps that were helpful in determining the missing residues.

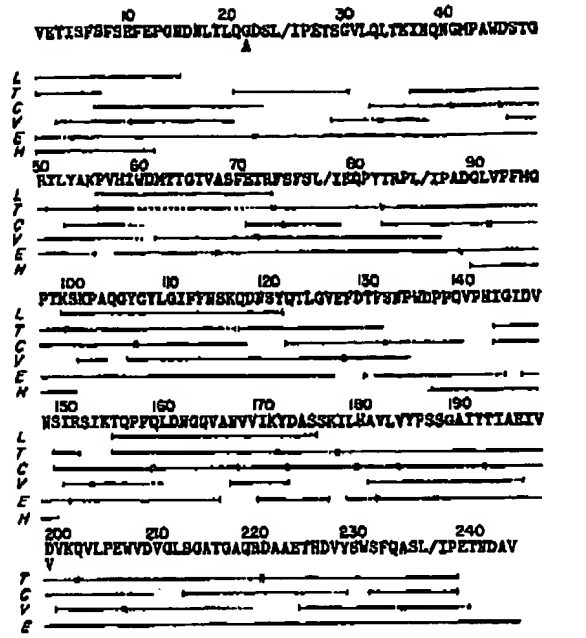


Fig.1. Alignment of ECorL and peptides used for sequence determination. L, peptides obtained from digestion with lysylendopeptidase-C; T, tryptic peptides; C, chymotryptic peptides; E and V, peptides obtained by digestion with elastase and *S. aureus* V8 protease, respectively; H, peptides obtained by hydrolysis with dilute formic acid in 6 M guanidine hydrochloride. Solid lines indicate regions of peptides sequenced by the manual DABITC/PITC method (except for the lysylendopeptidase-C peptides which were sequenced by the automatic sequencer). Dashed lines indicate residues which were not sequenced or yielded an unsatisfactory result.

Cleavage with dilute formic acid in 6 M guanidine HCl yielded 3 main fragments; one of them, resulting from hydrolysis of peptide bond Asp[136]-Pro[137], facilitated the elucidation of the sequence from amino acid 137 onwards. Each of the residues shown in fig.1 was identified at least twice in the sequence analysis using the manual DABITC/PITC double-coupling method or the automatic sequencer.

Microheterogenicity was observed in positions 22 (with Ala replacing Gly) and 199 (with Val replacing Asp), suggesting differences in the sequence of the two subunits of ECorL in these positions.

3.2. *Glycosylation site*

Digestion of ECorL by pronase, followed by gel filtration on Sephadex G-50, afforded a crude glycopeptide preparation in good yield (77% based on the neutral sugar content of the starting material). Fractionation of this preparation on reverse-phase HPLC led to the isolation of 5 main (glyco)peptides (fig.2A). HPLC of the crude glycopeptide after deglycosylation by endoglycosidase F (fig.2B), showed a single change in the enzyme-treated sample, i.e. a shift of peptide no.2 to a new position (peptide no.6),

3.4. *Homologies*

Comparison of the amino acid sequence of ECorL with that of other legume lectins (fig.3) reveals a high degree of homology. The amino acids in 39 positions are invariant in all the lectins listed. It is notable that these highly conserved amino acids include 4 which correspond to residues previously identified in ConA [17] as being important in the binding of Ca^{2+} and Mn^{2+},

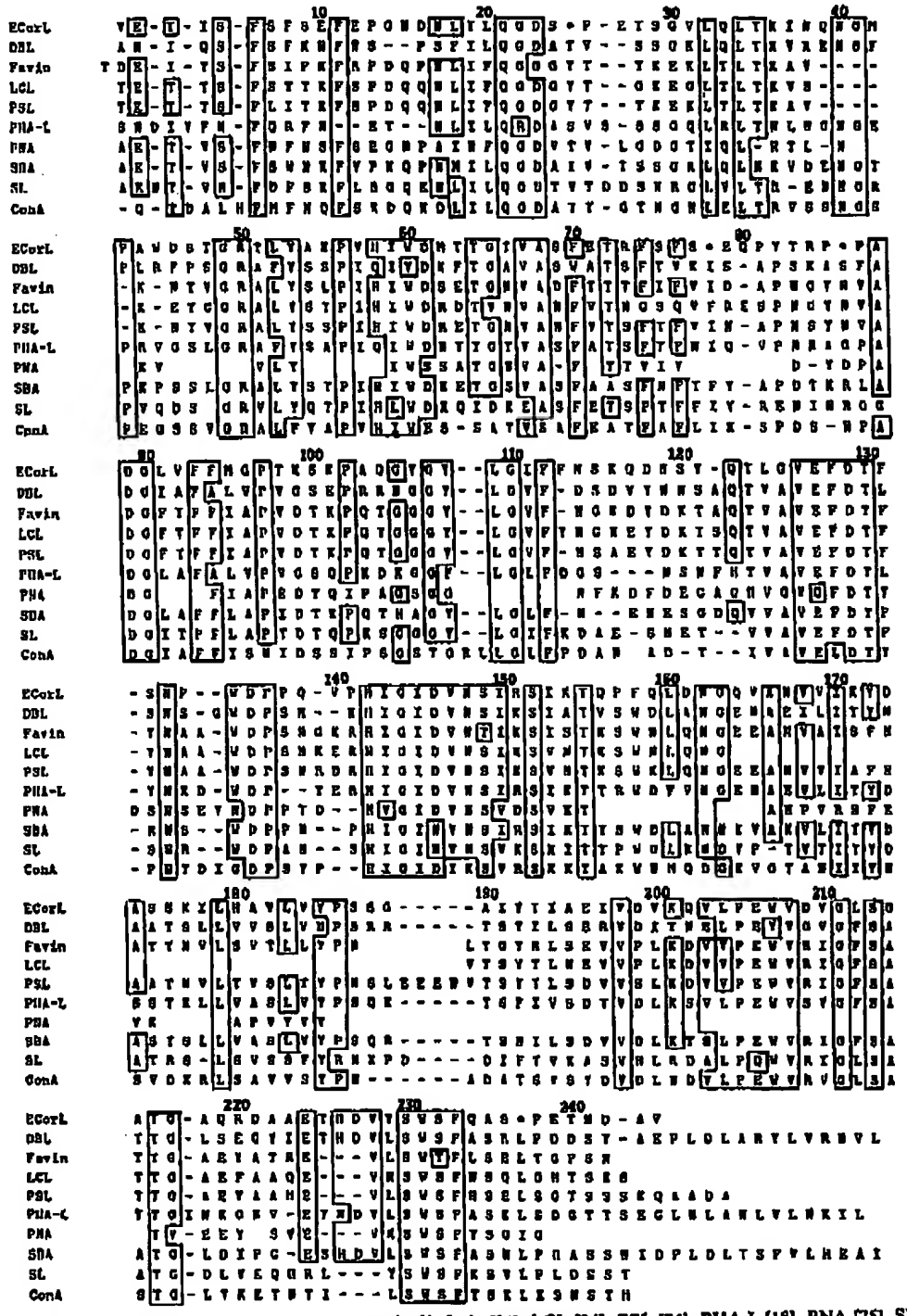


Fig.3. Comparison of amino acid sequences of ECorL, DBL [22,23], favin [24], LCL [24], PSL [24], PHA-L [18], PNA [25], SBA [24] and SL [24]. The sequence marked as ConA is that of pro-ConA, which exhibits circular homology with the mature lectin [26,27]. The identical residues have been boxed. –, Deletions inserted in all sequences to maximize homology; *, leucine or isoleucine.

i.e. $Asp^{10(129)}$, $Asp^{19(136)}$, $His^{24(142)}$, and $Ser^{34(152)}$. (The numbers without parentheses refer to the ConA sequence; those in parentheses are of pro-ConA after alignment with ECorL as given in fig.3.) Another metal binding residue, $Glu^{8(127)}$, is nearly invariant, except for peanut agglutinin in which glutamine was found instead. Similarly, the amino acids comprising the 3-dimensional structure of the hydrophobic cavity of ConA [17,18] are invariant ($Val^{89(208)}$, $Phe^{111(233)}$ and $Phe^{212(93)}$ or highly conserved (e.g. $Ser^{113(235)}$) in homologous positions in all the lectins. On the other hand, the residues which constitute the monosaccharide binding site in ConA [17] appear to be poorly conserved, for instance, $Asp^{208(89)}$ is the only residue maintained in this site in the other lectins (except for leukoagglutinin from phaseolus vulgaris in which it is replaced by a valine residue). It is also of interest that *N*-glycosylation triplets (Asn-X-Ser/Thr) are absent from some of the lectins; when present, they are located at different positions in the primary structure and are not always glycosylated. Thus, on ECorL, there is a single occupied glycosylation site at Asn^{17}; in PHA-L there are two occupied sites at positions 10 and 62 [18,19] and in soybean agglutinin there are 3 glycosylation sites at positions 40, 75 and 114, only one of which (position 75) is occupied [20]. In favin, the sugar moiety is located at Asn^{173} [21].

The similarity in the primary structure between ECorL and other legume lectins listed in fig.3 further supports the proposal that all these lectins share a common evolutionary origin. Their structures have been highly conserved in evolution, presumably for ensuring the maintenance of an important physiological function(s) yet to be determined.

Acknowledgements: We wish to thank Ms S. Haalman for the preparation of the lectin and Ms R. Samuel for valuable secretarial help. R.A. wishes to acknowledge a FEBS Fellowship for a short-term visit in the laboratory of Dr Michael Richardson. This work was supported by the US-Israel Binational Science Foundation Grant No.3727.

REFERENCES

[1] Iglesias, J.L., Lis, H. and Sharon, N. (1982) Eur. J. Biochem. 123, 247–252.

[2] Lis, H., Joubert, F.J. and Sharon, N. (1985) Phytochemistry 24, 2803–2809.

[3] Ashford, D., Dwek, R.A., Welply, J.K., Amatayakul, S., Homans, S.W., Lis, H., Taylor, G.N., Sharon, N. and Rademacher, T. (1987) Eur. J. Biochem. 166, 311–320.

[4] Foriers, A., Wuhrmart, C., Sharon, N. and Strosberg, D. (1977) Biochem. Biophys. Res. Commun. 4, 980–986.

[5] Lis, H. and Sharon, N. (1978) J. Biol. Chem. 253, 3468–3476.

[6] Richardson, M. and Granum, P.E. (1985) FEBS Lett. 182, 479–484.

[7] Fling, S.P. and Gregerson, D.S. (1986) Anal. Biochem. 155, 83–88.

[8] Matsudaira, P. (1987) J. Biol. Chem. 262, 10035–10038.

[9] Richardson, M., Campos, F.A.P., Moriera, R.A., Ainouz, I.L., Begbie, R., Watt, M.B. and Pusztai, A. (1984) Eur. J. Biochem. 144, 101–111.

[10] Yarwood, A., Richardson, M., Sousa-Cavada, B. and Rouge, P. (1985) FEBS Lett. 184, 104–109.

[11] Richardson, M., Yarwood, A. and Rouge, P. (1987) FEBS Lett. 216, 145–150.

[12] Dubois, M., Gilles, K.A., Hamilton, J.K., Rebers, P.A. and Smith, F. (1956) Anal. Chem. 28, 350–356.

[13] Chang, J.Y., Creaser, E.H. and Hughes, G.J. (1977) J. Chromatogr. 140, 125–128.

[14] Chang, J.Y., Brauer, D. and Wittmann-Liebold, B. (1978) FEBS Lett. 93, 205–214.

[15] Wellinder, K.G. (1976) FEBS Lett. 72, 19–23.

[16] Miller, J.B., Hsu, R., Heinrikson, R. and Yachnin, S. (1975) Proc. Natl. Acad. Sci. USA 72, 1388–1391.

[17] Becker, J.W., Reeke, G.N., Wang, J.L., Cunningham, B.A. and Edelman, G.M. (1975) J. Biol. Chem. 250, 1513–1524.

[18] Hoffman, L.M. and Donaldson, D.D. (1985) EMBO J. 4, 883–889.

[19] Sturm, A. and Chrispeels, M.J. (1986) Plant Physiol. 81, 320–322.

[20] Becker, J.W., Cunningham, B.A. and Hemperly, J.J. (1983) in: Chemical Taxonomy, Molecular Biology and Function of Plant Lectins (Goldstein, I.J. and Etzler, M.E. eds) pp.31–45, A.R. Liss, New York.

[21] Cunningham, B.A., Hemperly, J.J., Hopp, T.P. and Edelman, G.M. (1979) Proc. Natl. Acad. Sci. USA 76, 3218–3222.

[22] Schnell, D.J. and Etzler, M.E. (1987) J. Biol. Chem. 262, 7220–7225.

[23] Schnell, D.J. and Etzler, M.E. (1988) J. Biol. Chem. 263, 14648–14653.

[24] Strosberg, A.D., Buffard, D., Lauwereys, M. and Foriers, A. (1986) in: The Lectins: Properties, Functions and Applications in Biology and Medicine (Liener, I.E. et al. eds) pp.249–264, Academic Press.

[25] Lauwereys, M., Foriers, A., Sharon, N. and Strosberg, A.D. (1985) FEBS Lett. 181, 241–244.

[26] Hemperly, J.J. and Cunningham, B.A. (1983) Trends Biochem. Sci. 8, 100–102.

[27] Bowles, D.J., Marcus, S.E., Pappin, D.J.C., Findlay, J.B.C., Eliopoulos, E., Maycox, P.R. and Burgess, J. (1986) J. Cell Biol. 102, 1284–1297.

Structural analysis of two crystal forms of lentil lectin at 1.8 A resolution.

Loris R, Van Overberge D, Dao-Thi MH, Poortmans F, Maene N, Wyns L.

Laboratorium voor Ultrastructuur, Vrije Universiteit Brussel, Sint-Genesius-Rode, Belgium.

The structures of two crystal forms of lentil lectin are determined and refined at high resolution. Orthorhombic lentil lectin is refined at 1.80 A resolution to an R-factor of 0.184 and monoclinic lentil lectin at 1.75 A resolution to an R-factor of 0.175. These two structures are compared to each other and to the other available legume lectin structures. The monosaccharide binding pocket of each lectin monomer contains a tightly bound phosphate ion. This phosphate makes hydrogen bonding contacts with Asp-81 beta, Gly-99 beta, and Asn-125 beta, three residues that are highly conserved in most of the known legume lectin sequences and essential for monosaccharide recognition in all legume lectin crystal structures described thus far. A detailed analysis of the composition and properties of the hydrophobic contact network and hydrophobic nuclei in lentil lectin is presented. Contact map calculations reveal that dense clusters of nonpolar as well as polar side chains play a major role in secondary structure packing. This is illustrated by a large cluster of 24 mainly hydrophobic amino acids that is responsible for the majority of packing interactions between the two beta-sheets. Another series of four smaller and less hydrophobic clusters is found to mediate the packing of a number of loop structures upon the front sheet. A very dense, but not very conserved cluster is found to stabilize the transition metal binding site. The highly conserved and invariant nonpolar residues are distributed asymmetrically over the protein.

PMID: 7731952 [PubMed - indexed for MEDLINE]

Erratum in:

- Biochem Biophys Res Commun 2000 Apr 2;270(1):329.

Mode of molecular recognition of L-fucose by fucose-binding legume lectins.

Thomas CJ, Surolia A.

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, 560 012, India.

Recognition of cell surface carbohydrate moieties by lectins plays a vital role in many a biological process. Fucosyated residues are often implicated as key recognition markers in many cellular processes. In particular, the aspects of molecular recognition of fucose by fucose-bindinglectins UEA 1 and LTA pose a special case because no crystal structure of these lectins is available. The study was conducted to elucidate the process of recognition of l-fucose by UEA1 and LTA by correlating structure-based sequence alignment and other available biochemical/biophysical data. The study points out that the mode of recognition of l-fucose is coordinated by the invariant triad of residues the asparagine 137, glycine 105, and aspartate 87. The major hydrophobic stacking residue in this case is the tyrosine 220. The study also reiterates the key role of the conserved triad of residues in the combining site which is a common feature for all legume lectins whose crystal structures are known. Copyright 2000 Academic Press.

PMID: 10679191 [PubMed - indexed for MEDLINE]

About Entrez
NCBI Toolbar
Text Version
Entrez PubMed
Overview
Help | FAQ
Tutorials
New/Noteworthy
E-Utilities
PubMed Services
Journals Database
MeSH Database
Single Citation Matcher
Batch Citation Matcher
Clinical Queries
Special Queries
LinkOut
My NCBI
Related Resources
Order Documents
NLM Mobile
NLM Catalog
NLM Gateway
TOXNET
Consumer Health
Clinical Alerts
ClinicalTrials.gov
PubMed Central

1: J Biomol Struct Dyn. 1998 Apr;15(5):853-60. Related Articles, Links

Three dimensional structure of the soybean agglutinin Gal/GalNAc complexes by homology modeling.

Rao VS, Lam K, Qasba PK.

Structural Glycobiology Section, Laboratory of Experimental and Computational Biology, National Cancer Institute, NCI-FCRDC, Frederick, Maryland 21702, USA.

Complexes of soybean agglutinin (SBA) with galactose (Gal) and N-acetyl galactosamine (GalNAc) have been modeled based on its homology to erythrina corallodendron (EcorL) lectin. The three dimensional structure of SBA-Gal modeled with homology techniques agrees well with SBA-(beta-LacNAc)2Gal-R complex determined by X-ray crystallographic techniques at the beta-sheet regions and the regions where Ca2+ and Mn2+ ions bind. However, significant deviations have been observed between the modeled and the X-ray structures, particularly at the loop regions where the polypeptide chain could not be unequivocally traced in the X-ray structure. The hydrogen bonding scheme, predicted from the homology model, shows that the invariant residues i.e. Asp, Gly, Asn, and aromatic residues (Phe) found in all other legume lectins, bind Gal, slightly in a different way than reported in X-ray structure of SBA-pentasaccharide complex. The higher binding affinity of GalNAc over Gal to SBA is due to additional hydrophobic interactions with Tyr107 rather than a hydrogen bond between N-acetamide group of the sugar and the side chain of Asp88 as suggested from X-ray crystal structure studies. Our modeling also suggest that the variation in the length of the loop D observed among galactose binding legume lectins may not have any effect on the binding of sugar at the monosaccharide specific site of the lectins. Soybean agglutinin (SBA) is a member of the leguminous family of lectins. They generally possess a single carbohydrate binding site, besides the tightly bound Ca2+ and Mn2+ ions which are required for their carbohydrate binding activity. They possess a high degree of sequence homology and about 50% of the amino acid residues are invariant. Some of these invariant amino acid residues are involved in the binding of sugar moieties and in metal ion coordination. X-ray crystallographic studies showed that their three-dimensional structures are very similar, though they differ in their carbohydrate binding specificity (1-6). Three of the invariant residues Asp, Gly, and Asn, besides an aromatic residue (Phe or Tyr), are

involved in carbohydrate binding. Independent of their sugar specificity, these four residues in legume lectins provide the basic frame for the sugar to bind.

PMID: 9619508 [PubMed - indexed for MEDLINE]

1: Int J Biol Macromol. 1998 Nov;23(4):295-307.

Architecture of the sugar binding sites in carbohydrate binding proteins--a computer modeling study.

Rao VS, Lam K, Qasba PK.

Structural Glycobiology Section, Laboratory of Experimental and Computational Biology, National Cancer Institute, NCI-FCRDC, Frederick, MD 21702-1201, USA.

Different sugars, Gal, GalNAc and Man were docked at the monosaccharide binding sites of Erythrina corallodenron (EcorL), peanut lectin (PNA), Lathyrus ochrus (LOLI), and pea lectin (PSL). To study the lectin-carbohydrate interactions, in the complexes, the hydroxymethyl group in Man and Gal favors, gg and gt conformations respectively, and is the dominant recognition determination. The monosaccharide binding site in lectins that are specific to Gal/GalNAc is wider due to the additional amino acid residues in loop D as compared to that in lectins specific to Man/Glc, and affects the hydrogen bonds of the sugar involving residues from loop D, but not its orientation in the binding site. The invariant amino acid residues Asp from loop A, and Asn and an aromatic residue (Phe or Tyr) in loop C provides the basic architecture to recognize the common features in C4 epimers. The invariant Gly in loop B together with one or two residues in the variable region of loop D/A holds the sugar tightly at both ends. Loss of any one of these hydrogen bonds leads to weak interaction. While the subtle variations in the sequence and conformation of peptide fragment that resulted due to the size and location of gaps present in amino acid sequence in the neighborhood of the sugar binding site of loop D/A seems to discriminate the binding of sugars which differ at C4 atom (galacto and gluco configurations). The variations at loop B are important in discriminating Gal and GalNAc binding. The present study thus provides a structural basis for the observed specificities of legume lectins which uses the same four invariant residues for binding. These studies also bring out the information that is important for the design/engineering of proteins with the desired carbohydrate specificity.

PMID: 9849627 [PubMed - indexed for MEDLINE]